GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2000, 18:23:48 ; Search time 983.13 Seconds
(without alignments)
1900.641 Million cell updates/sec

Title: US-09-257-585-1
Perfect score: 1047
Sequence: 1 gatctgtaaattacatatat.....atggagqcgacggqcggatcc 1047

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 972840 seqs, 892848109 residues

Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0
Maximum DB seq length: 100000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: qb_ba1:*
2: qb_ba2:*
3: qb_om:*
4: qb_ov:*
5: qb_pat:*
6: qb_ph:*
7: qb_pl1:*
8: qb_pl2:*
9: qb_pr1:*
10: qb_pr2:*
11: qb_pr3:*
12: qb_ro:*
13: qb_sts:*
14: qb_sy:*
15: qb_un:*
16: em_tun:*
17: em_hum1:*
18: em_hum2:*
19: em_in:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_sy:*
29: em_un:*
30: em_vi:*
31: qb_htq1:*
32: qb_htq2:*
33: qb_in1:*
34: qb_in2:*
35: em_ba1:*
36: em_ba2:*
37: em_hum3:*
38: em_hum4:*
39: qb_pr4:*
40: qb_htq3:*
41: qb_htq4:*
42: qb_htq5:*
43: qb_htq6:*
44: qb_htq7:*
45: em_htq1:*
46: em_htq2:*
47: em_htq3:*
48: em_hum5:*
49: qb_pr5:*
50: qb_pr6:*
51: qb_htq8:*
52: qb_htq9:*
53: qb_htq10:*
54: qb_htq11:*
55: qb_htq12:*
56: qb_htq13:*
57: qb_htq14:*
58: qb_in3:*
59: qb_htq15:*
60: qb_htq16:*
61: qb_htq17:*
62: em_htq4:*
63: em_htq5:*
64: em_htq6:*
65: em_htq7:*
66: em_hum6:*
67: qb_htq18:*
68: qb_htq19:*
69: qb_htq20:*
70: qb_htq21:*
71: qb_htq22:*
72: qb_htq23:*
73: qb_htq24:*
74: qb_htq25:*
75: qb_htq26:*
76: qb_htq27:*
77: qb_htq28:*
78: qb_htq29:*
79: qb_htq30:*
80: qb_htq31:*
81: qb_vi1:*
82: qb_vi2:*

Pred. No. is the number of results predicted by
score greater than or equal to the score of the
and is derived by analysis of the total score

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID
1	44.2	4.2	44	2	AC034140
2	44	4.2	4560	1	C10PDR
3	43.4	4.1	973	33	DSU13899
4	43	4.1	15049	2	AE001482
5	42.8	4.1	973	33	DSU132900
6	42.6	4.1	243747	34	AF165926
7	41.8	4.0	81542	1	AB026647
8	41.6	4.0	97208	44	AC007396
9	40.8	3.9	88041	43	AC022522
10	40.8	3.9	123389	8	F12F1
11	40.6	3.9	23418	50	AC004153
12	40.4	3.9	474	34	AF153647
13	40.4	3.9	480	34	AF010513
14	40	3.8	3439	33	TETMTRGTG
15	40	3.8	47299	34	AF160864
16	39.8	3.8	4500	2	SEU43366
17	39.6	3.8	969	33	DMA132902
18	39.6	3.8	11439	34	TELC01B4
19	39.6	3.8	67944	34	AC006722
20	39.6	3.8	150495	41	AC006732
21	39.6	3.8	206217	41	AC006754
22	39.4	3.8	172151	37	AC023672
23	39.2	3.7	152604	42	CNS01DXE
24	39.2	3.7	218221	41	PFMAL13P3

	25	39.2	3.7	349919	54	AC008576	AC008576 Homo sapi
	26	39	3.7	55232	76	AC055771	AC055771 Homo sapi
c	27	39	3.7	68270	33	CEY7A5A	AL034489 Caenorhab
	28	39	3.7	86829	33	PFMAL3P5	AL034556 Plasmodiu
	29	39	3.7	211038	32	CNS01D20	AL121829 Homo sapi
c	30	39	3.7	235922	31	CEY7A5	AL021576 Caenorhab
	31	39	3.7	261927	32	HSA151N12	AL109835 Homo sapi
c	32	38.8	3.7	1163	33	AF044861	AF044861 Eurema ni
c	33	38.8	3.7	74229	33	CEY111B2D	AL132885 Caenorhab
c	34	38.8	3.7	110000	32	CEY111B2_0	Z98857 Caenorhabdi
c	35	38.8	3.7	142737	49	AC027658	AC027658 Oryza sat
c	36	38.8	3.7	149594	72	AC021747	AC021747 Homo sapi
c	37	38.8	3.7	150613	7	AB001684	AB001684 Chlorella
c	38	38.8	3.7	153670	60	AC016098	AC016098 Homo sapi
c	39	38.8	3.7	179871	31	AP001400	AP001400 Homo sapi
c	40	38.8	3.7	230601	41	AC010813	AC010813 Homo sapi
c	41	38.6	3.7	1154	33	AF044864	AF044864 Pieris ra
	42	38.6	3.7	4399	2	BBU35673	U35673 Borrelia bu
c	43	38.6	3.7	14239	2	AE001133	AE001133 Borrelia
c	44	38.6	3.7	26798	42	AC017465	AC017465 Drosophil
c	45	38.6	3.7	33539	34	CELF54E7	U00067 Caenorhabdi

ALIGNMENTS

RESULT 1
AC034140/c
LOCUS AC034140 44102 bp DNA HTG 04-APR-2000
DEFINITION Homo sapiens clone RP11-102N11, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC034140
VERSION AC034140.1 GI:7408089
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 44102)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 44102)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (04-APR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT Center project name: H_NH0102N11.
* NOTE: This record contains 89 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 520: contig of 520 bp in length
* 521 530: gap of unknown length
* 531 1026: contig of 496 bp in length
* 1027 1036: gap of unknown length
* 1037 1536: contig of 500 bp in length
* 1537 1546: gap of unknown length
* 1547 1690: contig of 144 bp in length
* 1691 1700: gap of unknown length
* 1701 2218: contig of 518 bp in length
* 2219 2228: gap of unknown length
* 2229 2742: contig of 514 bp in length
* 2743 2752: gap of unknown length
* 2753 3194: contig of 442 bp in length
* 3195 3204: gap of unknown length
* 3205 3255: contig of 51 bp in length
* 3256 3265: gap of unknown length
* 3266 3777: contig of 512 bp in length
* 3778 3787: gap of unknown length
* 3788 4299: contig of 512 bp in length
* 4300 4309: gap of unknown length
* 4310 4829: contig of 520 bp in length
* 4830 4839: gap of unknown length
* 4840 5358: contig of 519 bp in length
* 5359 5368: gap of unknown length
* 5369 5888: contig of 520 bp in length
* 5889 5898: gap of unknown length
* 5899 6242: contig of 344 bp in length
* 6243 6252: gap of unknown length
* 6253 6770: contig of 518 bp in length
* 6771 6780: gap of unknown length
* 6781 7259: contig of 479 bp in length
* 7260 7269: gap of unknown length
* 7270 7781: contig of 512 bp in length
* 7782 7791: gap of unknown length
* 7792 8119: contig of 328 bp in length
* 8120 8129: gap of unknown length
* 8130 8409: contig of 280 bp in length
* 8410 8419: gap of unknown length
* 8420 8915: contig of 496 bp in length
* 8916 8925: gap of unknown length
* 8926 9437: contig of 512 bp in length
* 9438 9447: gap of unknown length
* 9448 9634: contig of 187 bp in length
* 9645 9644: gap of unknown length
* 9645 10164: contig of 520 bp in length
* 10165 10174: gap of unknown length
* 10175 10693: contig of 519 bp in length
* 10694 10703: gap of unknown length
* 10704 11182: contig of 479 bp in length
* 11183 11192: gap of unknown length
* 11193 11710: contig of 518 bp in length
* 11711 11720: gap of unknown length
* 11721 12238: contig of 518 bp in length
* 12239 12248: gap of unknown length
* 12249 12760: contig of 512 bp in length
* 12761 12770: gap of unknown length
* 12771 13283: contig of 513 bp in length
* 13284 13293: gap of unknown length
* 13294 13804: contig of 511 bp in length
* 13805 13814: gap of unknown length
* 13815 14326: contig of 512 bp in length
* 14327 14336: gap of unknown length
* 14337 14840: contig of 504 bp in length
* 14841 14850: gap of unknown length
* 14851 15362: contig of 512 bp in length
* 15363 15372: gap of unknown length
* 15373 15891: contig of 519 bp in length
* 15892 15901: gap of unknown length
* 15902 16670: contig of 769 bp in length
* 16671 16680: gap of unknown length
* 16681 17128: contig of 448 bp in length
* 17129 17148: gap of unknown length
* 17139 17658: contig of 520 bp in length
* 17659 17668: gap of unknown length
* 17669 17719: contig of 51 bp in length
* 17720 17729: gap of unknown length
* 17730 18145: contig of 416 bp in length
* 18146 18155: gap of unknown length
* 18156 18661: contig of 506 bp in length
* 18662 18671: gap of unknown length
* 18672 19183: contig of 512 bp in length
* 19184 19193: gap of unknown length
* 19194 19593: contig of 400 bp in length
* 19594 19603: gap of unknown length
* 19604 20113: contig of 510 bp in length
* 20114 20123: gap of unknown length
* 20124 20880: contig of 757 bp in length

```
*    20881    20990: gap of unknown length
*    20891    21248: contig of 358 bp in length
*    21249    21258: gap of unknown length
*    21259    21776: contig of 518 bp in length
*    21777    21786: gap of unknown length
*    21787    22154: contig of 368 bp in length
*    22155    22164: gap of unknown length
*    22165    22620: contig of 456 bp in length
*    22621    22630: gap of unknown length
*    22631    23149: contig of 519 bp in length
*    23150    23159: gap of unknown length
*    23160    23510: contig of 351 bp in length
*    23511    23520: gap of unknown length
*    23521    24031: contig of 511 bp in length
*    24032    24041: gap of unknown length
*    24042    24552: contig of 511 bp in length
*    24553    24562: gap of unknown length
*    24563    25074: contig of 512 bp in length
*    25075    25084: gap of unknown length
*    25085    25548: contig of 464 bp in length
*    25549    25558: gap of unknown length
*    25559    26070: contig of 512 bp in length
*    26071    26080: gap of unknown length
*    26081    26597: contig of 517 bp in length
*    26598    26607: gap of unknown length
*    26608    27095: contig of 488 bp in length
*    27096    27105: gap of unknown length
*    27106    27624: contig of 519 bp in length
*    27625    27634: gap of unknown length
*    27635    28144: contig of 510 bp in length
*    28145    28154: gap of unknown length
*    28155    28675: contig of 521 bp in length
*    28676    28685: gap of unknown length
*    28686    29204: contig of 519 bp in length
*    29205    29214: gap of unknown length
*    29215    29726: contig of 512 bp in length
*    29727    29736: gap of unknown length
*    29737    30248: contig of 512 bp in length
*    30249    30258: gap of unknown length
*    30259    30770: contig of 512 bp in length
*    30771    30780: gap of unknown length
*    30781    31228: contig of 448 bp in length
*    31229    31238: gap of unknown length
*    31239    31757: contig of 519 bp in length
*    31758    31767: gap of unknown length
*    31768    32286: contig of 519 bp in length
*    32287    32296: gap of unknown length
*    32297    32815: contig of 519 bp in length
*    32816    32825: gap of unknown length
*    32826    33344: contig of 519 bp in length
*    33345    33354: gap of unknown length
*    33355    33872: contig of 518 bp in length
*    33873    33882: gap of unknown length
*    33883    34400: contig of 518 bp in length
*    34401    34410: gap of unknown length
*    34411    34906: contig of 496 bp in length
*    34907    34916: gap of unknown length
*    34917    35324: contig of 408 bp in length
*    35325    35334: gap of unknown length
*    35335    35846: contig of 512 bp in length
*    35847    35856: gap of unknown length
*    35857    36352: contig of 496 bp in length
*    36353    36362: gap of unknown length
*    36363    36810: contig of 448 bp in length
*    36811    36820: gap of unknown length
*    36821    37335: contig of 515 bp in length
*    37336    37345: gap of unknown length
*    37346    37864: contig of 519 bp in length
*    37865    37874: gap of unknown length
*    37875    38394: contig of 520 bp in length
*    38395    38404: gap of unknown length
*    38405    38916: contig of 512 bp in length
*    38917    38926: gap of unknown length
*    38927    39439: contig of 513 bp in length
*    39440    39449: gap of unknown length
*    39450    40329: contig of 880 bp in length
*    40330    40339: gap of unknown length
*    40340    40858: contig of 519 bp in length
*    40859    40868: gap of unknown length
*    40869    41348: contig of 480 bp in length
*    41349    41358: gap of unknown length
*    41359    41870: contig of 512 bp in length

  Query Match              4.4%;  Score 44.2;  DB 72:
  Best Local Similarity    52.4%;  Pred. No. 1.5;
  Matches   97;  Conservative    ;;  Mismatches   88;

Qy       7 taaattacatatatttatttaccaqcatttqctaaccqcctatt
           ||  ||| | |||| |  ||        ||    | |    | |
Db   37294 TATTTTATAAATATATTATTTTAATATAAATATCATTTATTCTT

Qy      67 acgattaaataaccqcaattcattttqcqttctacacaaattaacc
           | ||||| |   | |  .  .  ' '  |  ||   ||||
Db   37254 AGTATTATGTTAGTATCTTCATATTATAGAATTACTAGAATTATAT

Qy     127 qtatttaaqcqaaaatttqccattttttaatccattttaaaa
           ||   ||   | |   |||.|||.|   |  |   || |||
Db   37174 CTAAAAAAAAAATGCATTGGCATTTTTTACTTGCATGCCTAAAA

Qy     187 accca 191
             |||
Db   37114 GACCA 37110


RESULT   2
CLOPFOR/c
LOCUS       CLOPFOR      4560 bp    DNA              B
DEFINITION  Clostridium perfringens pfoR gene, complet
ACCESSION   M81080 M34727
VERSION     M81080.1  GI:144886
KEYWORDS    perfringolysin O; thiol activated hemolys
SOURCE      Clostridium perfringens (library: lambda
  ORGANISM  Clostridium perfringens
            Bacteria; Firmicutes; Bacillus/Clostridi
            Clostridium.
REFERENCE   1  (bases 1 to 4560)
  AUTHORS   Shimizu,T., Okabe,A., Minami,J. and Haya
  TITLE     An upstream regulatory sequence stimulat
            perfringolysin O gene of Clostridium per
  JOURNAL   Infect. Immun. 59, 137-142 (1991)
  MEDLINE   91099951
FEATURES             Location/Qualifiers
     source          1..4560
                     /organism="Clostridium perfrin
                     /db_xref="taxon:1502"
                     /tissue_lib="lambda gt10"
     gene            2992..3404
                     /gene="pfoR"
     CDS             2992..3404
                     /gene="pfoR"
                     /codon_start=1
                     /transl_table=11
                     /product="perfringolysin O"
                     /protein_id="AAA23271.1"
                     /db_xref="GI:144886"
                     /translation="MIPEKKTKLIASIAMA
                     LSYNRNEVLAYNGDKTESFVPKESKKANKI
                     YISNLSLSMKNFVENKFTLLMVKLKFININ
                     IVSKWNEKYNSIHTLIAKIQYSKEMVYSKS
                     KKVMTLAYKQTEYIVSADLPKNMSDLETAV
                     IYVKLETISSSKDVQAAFKALIKNIPEKNS
                     IKDFDELKKVIKDNATFSTKNPAYPISYTS
                     KINLDHSGAYVAQFEVAWDEVSYLKEGNEV
                     NLSNAGEKLGLAWEWWKDVISEYDVFLIN"
BASE COUNT     1755 a    5 c    705 g    1593 t
ORIGIN
```

```
  Query Match             4.2%;  Score 44;  DB 1;  Length 4560;
  Best Local Similarity   52.8%;  Pred. No. 1.8;
  Matches   95;  Conservative    0;  Mismatches   85;  Indels    0,  Gaps    0;

Qy       5 tgtaaaattacatatatttattttqccgcattttgctaaccgcctattcttttaaaattt 64
           | | |||    || |||| |||       |||      ||||  | ||| ||||   | |||
Db    2552 TTTGAATCCACTAAATTTTTTTCTATTAATTAAAAAAACCTTCAATTTTTTTGTTATTTT 2493

Qy      65 taacgattaaagaacgcaattcaatttgcgttctacacaaattaacaattcgtccaaaaq 124
           ||   |  || ||      |  | ||    || |||  || ||||  |||      ||||
Db    2492 TATTAACAAAAAAGATTTTAAATTTTATTTTTTACTAAATTTAAAGATTAAATTAAAAA 2433

Qy     125 agttatttaaagcgaagattttgccattttttttaatccattttaaaaaaatacatctgctt 184
           |||  |  |    || || |    || |||| || |||   ||| |  |||  | |  ||
Db    2432 AGGGGTGCACCAAAATATATAAAGTTATTTTGATACATCCCTAATATTATATTTTTAATT 2373


RESULT   3
DSU132899
LOCUS       DSU132899     973 bp    DNA             INV       21-MAR-2000
DEFINITION  Drosophila subobscura mitochondrial A+T-rich region, isolate subR.
ACCESSION   AJ132899
VERSION     AJ132899.1  GI:4454152
KEYWORDS    A+T-rich region.
SOURCE      Drosophila subobscura.
  ORGANISM  Mitochondrion Drosophila subobscura
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 973)
  AUTHORS   Brehm,A., Hernandez,M., Cabrera,V.M., Larruga,J.M., Pinto,F.M. and
            Gonzalez,A.M.
  TITLE     Nucleotide Sequence Structure and Phylogeny of the Complete
            A+T-rich Region of Mitochondrial DNA in the Drosophila subobscura
            Subgroup
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 973)
  AUTHORS   Gonzalez,A.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-FEB-1999) Gonzalez A.M., Department of Genetics,
            Faculty of Biology, University of La Laguna, Tenerife, E 38271,
            SPAIN
FEATURES             Location/Qualifiers
     source          1. .973
                     /organism="Drosophila subobscura"
                     /organelle="mitochondrion"
                     /isolate="Canary Islands Tenerife subR"
                     /db_xref="taxon:7241"
                     /dev_stage "Adult"
     tRNA            <1. .10
                     /gene="tRNA-Ile"
                     /evidence=not_experimental
                     /product="transfer RNA-Ile"
     gene            1. .10
                     /gene="tRNA-Ile"
     misc_feature    11. .950
                     /note="A+T-rich region"
                     /evidence=not_experimental
     rRNA            951. .>973
                     /note="partial sequence"
                     /evidence=not_experimental
                     /product="small ribosomal RNA"
BASE COUNT      409 a     93 c     42 g    489 t
ORIGIN


  Query Match             4.1%;  Score 43.4;  DB 33;  Length 973;
  Best Local Similarity   53.9%;  Pred. No. 2.7;
  Matches   89;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy      22 tatttttgccgcattttgctaaccgcctattcttttttaaaattttaacgattaaagaacgca 81
           |||  |  ||   ||| | |  |    | || | |||||| | ||||| || | |
Db     811 TATGTCACCATTTTTGGATTGCGAAATTTTTTATTTAAATTGTTAACTATATAAACATAT 870

Qy      82 attcaattttgcgttctacacaaattaacaattcgtccaaaaqaagttatttaaagcgaaga 141
           ||| ||  ||    | || | | |||| | |||  || | ||   |||| |||   || |
Db     871 ATTAAACATTTATATATATATATATTATCTATTAATCTAGAACTAGTATATAAAAAAAAA 930

Qy     142 tttggcatttttttttaatccatttttaaaaaaatacatctgcttta 186
               |  ||||      || | || |||| |||   |  ||||
Db     931 AAAATAGTAATTTTTTAATATATATATAAAATTACTATTAATTTA 975


RESULT   4
AE001482/c
LOCUS       AE001482     15549 bp    DNA             BCT       20 JAN 1999
DEFINITION  Helicobacter pylori, strain J99 section 43 of 132 of the complete
            genome.
ACCESSION   AE001482 AE001439
VERSION     AE001482.1  GI:4155018
KEYWORDS    .
SOURCE      Helicobacter pylori J99.
  ORGANISM  Helicobacter pylori J99
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 15549)
  AUTHORS   Alm,R.A., Ling,L.-S.L., Moir,D.T., King,B.L., Brown,E.D.,
            Doig,P.C., Smith,D.R., Noonan,B., Guild,B.C., deJonge,B.L.,
            Carmel,G., Tummino,P.J., Caruso,A., Uria-Nickelsen,M., Mills,D.M.,
            Ives,C., Gibson,R., Merberg,D., Mills,S.D., Jiang,Q., Taylor,D.E.,
            Vovis,G.F. and Trust,T.J.
  TITLE     Genomic-sequence comparison of two unrelated isolates of the human
            gastric pathogen Helicobacter pylori
  JOURNAL   Nature 397 (6715), 176-180 (1999)
  MEDLINE   99120557
  REMARK    Erratum:[published erratum appears in Nature 1999 Feb
            25;397(6721):719]
REFERENCE   2  (bases 1 to 15549)
  AUTHORS   King,B.L., Alm,R.A. and Trust,T.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (12 JAN-1999) Astra Research Center Boston, 128 Sidney
            Street, Cambridge, MA 02139, USA
COMMENT     Address all correspondence to: hp@arcb.us.astra.com or Richard
            A. Alm, Astra Research Center Boston, 128 Sidney Street, Cambridge,
            MA, 02139. Lo-Ssu L. Ling, Donald T. Moir, Douglas R. Smith,
            Braydon C. Guild, Gilles Carmel, Anthony Caruso, Debra M. Mills,
            Rene Gibson, and Gerald F. Vovis are with Genome Therapeutics
            Corporation, 100 Beaver Street, Waltham, MA, 02453. Qin Jiang and
            Diane E. Taylor are with the University of Alberta Department of
            Medical Microbiology and Immunology, Edmonton, Alberta, Canada, T6G
            2H7 and the Canadian Bacterial Diseases Network. All other authors
            are with Astra Research Center Boston, 128 Sidney Street,
            Cambridge, MA, 02139. Putative identifications, sequence
            alignments, and name and sequence search capability are available
            at ARCB's World Wide Web site. (URL:
            http://www.astra-boston.com/hpylori).
FEATURES             Location/Qualifiers
     source          1. .15549
                     /organism="Helicobacter pylori J99"
                     /strain="J99"
                     /db_xref="taxon:85963"
     gene            222. .878
                     /gene="orf18"
                     /note="jhp0480"
     CDS             222. .878
                     /gene="orf18"
                     /note="similar to H. pylori 26695 gene HP0531"
                     /codon_start=1
                     /transl_table=11
                     /product="cag island protein"
                     /protein_id "AAD06068.1"
                     /db_xref "GI:4155029"
                     /translation="MNDTTEHHGSNPLNAPPPSNSQSNDLLNLLDSLYPKGSLGEQRF
                     HEALKNQEELKNILIEIEKLPQFKRYFTLMQIQQAKQRIMEAYAHSFLGYIGGLEHLL
```

```
                GLCMGGIFVLFALYFVFLKTSKNTELVESLKTKFLKDQYFYYAEGVCAVLFPCLFLIPS
                IYELYILGIGSTNCFVLFVLFNTCFLIMGYLIYFVIFVIC KNFINGLLASLFCAF
                "
gene            914. .1756
                /gene="caqI"
                /note "jhp0481"
CDS             914. .1756
                /gene "caqI"
                /note="similar to H. pylori 26695 gene HP0532"
                /codon_start=1
                /transl_table=11
                /product="caq island protein"
                /protein_id="AAD06069.1"
                /db_xref="GI:4155030"
                /translation="MFLRASVFIGVACFCLIISAISNYAKKVVKQFNHVYTFVVNFLI
                EKYSIGLNDKFKGFFPMVQVKLFNYKFYLLGNSVVFFSFLVFHSKNLIGSDANKI
                LQYKNYFQANGARSDIDFYIQPILNQKGVVMIASNYNQNPNNKFKPQTFDVIQGSQPM
                LGANTKNIHGYDVSGANNKQVINFVAPFKAQLFKINQYYFLIQDKFQFVTIFKNNQS
                EILELSLSNRAGVQMPQNVISSFIFKNQNINMQAKFFFVPFKIQPFFFNFYLFNQIRCL
                LSGK"
gene            complement(1961. .2560)
                /gene "caqS"
                /note="jhp0482"
CDS             complement(1961. .2560)
                /gene="caqS"
                /note "similar to H. pylori 26695 gene HP0534"
                /codon_start=1
                /transl_table=11
                /product "caq island protein"
                /protein_id="AAD06070.1"
                /db_xref="GI:4155031"
                /translation="MSNNMPKLFSMIANSKDKKFKLIFSIQFNELINTDPKKKIIDQI
                KIMHDFKAQMHLNKGAIIKVFENYMKDYRAVIKSICVDKFKKVYRLLESETMELIHAI
                AENPNFLFSKFDKSILGIFLPFFSKPIMKNMSIKFMDSVIELYCINLFPLKLFVMFDE
                EVNFYANLKIIKQYNDYVRDLLMKFDLEKYMEEKGVQNA"
gene            complement(2999. .3379)
                /gene="caqQ"
                /note="jhp0483"
CDS             complement(2999. .3379)
                /gene "caqQ"
                /note="similar to H. pylori 26695 gene HP0535"
                /codon_start=1
                /transl_table=11
                /product="caq island protein"
                /protein_id="AAD06071.1"
                /db_xref="GI:4155032"
                /translation="MKIFKISIKEGNKLIIQVFLKGLSMKFKIASLGGGINGFTGI
                SSDLFSMILNFLFFLMLMMGINFALGKKFNIPMDNIKNFMAEVLKNGFDSIKNMGSAL
                VGNGFGSKSDKITNKMSVPQVRL"
gene            complement(3806. .4150)
                /gene="caqP"
                /note="jhp0484"
CDS             complement(3806. .4150)
                /gene "caqP"
                /note="similar to H. pylori 26695 gene HP0536"
                /codon_start=1
                /transl_table=11
                /product="caq island protein"
                /protein_id="AAD06072.1"
                /db_xref="GI:4155033"
                /translation="MKRPISKLKQNFIQFKHSFNKRLDKYSLYYRLFNISSIVIGFLI
                ALFSYGAGVILVYPILFLFALIIKPSFFYYTTYLLLLVSLSIISKYYLLSHANFTMKL
                IMLMIQWQNWFL"
gene            4765. .5895
                /gene="caqM"
                /note="jhp0485"
CDS             4765. .5895
                /gene="caqM"
                /note "similar to H. pylori 26695 gene HP0537"
                /codon_start=1
                /transl_table=11
                /product="caq island protein"
                /protein_id="AAD06058.1"
                /db_xref "GI:4155019"
                /translation "MLAGIVTSGLVAI[illegible]
                KAKKEIENNSHFIIIVEKQKIELVEVRAKVI
                IAKSEKYIIIGAKIKNIIKKELLPNFYAYVINQI
                KKIAIGKKAGAIIGMGVESDLFQNKKASKIAI
                NIIYVLIIENKTIGKAKIIAGKATNILIFNIDS
                LKNVKIIVILNIIKNFINQKLSFAGKQKLCLLV
                VGGIFMLIFIFMGKIIKIIFNFTKFVVSGVINN
gene            5910. .6845
                /gene "caqN"
                /note "jhp0486"
CDS             5910. .6845
                /gene "caqN"
                /note "similar to H. pylori 2669
                /codon_start=1
                /transl_table=11
                /product "caq island protein"
                /protein_id="AAD06059.1"
                /db_xref="GI:4155020"
                /translation "MKSIFKKLGSVALYSLVVG
                IKELVA
                IKIFYGFHIFIIIKIFKKEFAINALFKELFNDKA
                AKHKGLGLNGINANILNLNFNGNSSILIKSNKI
                LWKIIIIIINFGLSSNDYDGLKKGLIKGIKI
                HLSGLKLALNNSLNKIFLIASGAIGKKEKEME
                DATPSNNKSKGELLNIVINQIEQSIAHGKK"
gene            complement(6912. .7625)
                /gene "caqI"
                /note "jhp0487"
CDS             complement(6912. .7625)
                /gene="caqI"
                /note "similar to H. pylori 2669
                /codon_start=1
                /transl_table=11
                /product="caq island protein"
                /protein_id "AAD06060.1"
                /db_xref "GI:4155021"
                /translation "MKTLVKNTIYSFLLLSVLL
                LDFITSTTSNSANNKIGQEDALNIKKAAIALK
                YMSSFFLLIYMKINFIQKIAFQQCGISNKV
                MKQSKIAFQIVGLSNNSIK
                ISQKKLYIIKSQSSIIVK"
gene            complement(7622. .8767)
                /gene "caqI"
                /note "jhp0488"
CDS             complement(7622. .8767)
                /gene "caqI"
                /note "similar to H. pylori 2669
                /codon_start=1
                /transl_table=11
                /product "caq island protein"
                /protein_id "AAD06061.1"
                /db_xref "GI:4155022"
                /translation "MKCFLSIFSFLIFCGLS
                AQAFLKAIEAQSNAKEKAIQAQIEGELRIQIA
                GIIIVMFIIGISAIGSNVFMIVEQKIN
                NAIIFALKNSFFPSVTLKTQMEMLSQALGSSS
                IFIAMLFAIIIYNSLTCQALFQSIIYAPALS
                GLKIIAGLMLQVIQNYKMIVALIQNKLSDAI
                EVAIINAENGKNYIENLEKKADLWEEQIKIF
gene            complement(8778. .9890)
                /gene="caqH"
                /note="jhp0489"
CDS             complement(8778. .9890)
                /gene "caqH"
                /note "similar to H. pylori 266
                /codon_start=1
                /transl_table=11
                /product "caq island protein"

Query Match             4.1%;  Score 43;  DB 2;
Best Local Similarity   5.9%;  Pred. No. 1;
Matches  100;  Conservative    0;  Mismatches   9;
```

```
Qy       1 gatctgtaaattacatatatttatttttgccgcatttttgctaaccgccctattctttttaaa 60
           | | | | || |   ||| ||  || |  ||||||||   |||||    | ||||||||||
Db    2871 GTTATCTTAACTCTTTATTTTAGTTGTAACGCATTTTTACAACCGATCTTGCTTTTTTAAA 2812

Qy      61 attttaacgattaagaacgcaattcaatttgcgttctacacaaattaacaattcgtccca 120
           | |      | ||  ||  |  || || |  |  ||     | | ||   ||  |||
Db    2811 AATAATTATAATATTAAAACTTTTTAAATCGGTTTTTAGGGCTATTTGTGAAAGCGTTTT 2752

Qy     121 aaaaagctatttaagcgaagatttgcgatttttttaatccatttttaaaaaaatacatct 180
           |||| || ||   |||| | | |  || || | || ||| |     | | |
Db    2751 TTGTAGCTTTTGAAAAATACATATAACAGTGATTGAAGCAATATTTTATGGTTAAAAAAT 2692

Qy     181 gctttaaccgacctt 195
           | |||    | | ||
Db    2691 GGTTTCTTGCTCTTT 2677


RESULT   5
DSU132900
LOCUS       DSU132900      973 bp    DNA             INV       21-MAR-2000
DEFINITION  Drosophila subobscura mitochondrial A+T-rich region, isolate subl.
ACCESSION   AJ132900
VERSION     AJ132900.1  GI:4454153
KEYWORDS    AT-rich region.
SOURCE      Drosophila subobscura.
  ORGANISM  Mitochondrion Drosophila subobscura
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 973)
  AUTHORS   Brehm,A., Hernandez,M., Cabrera,V.M., Larruga,J.M., Pinto,F.M. and
            Gonzalez,A.M.
  TITLE     Nucleotide Sequence Structure and Phylogeny of the Complete
            A+T rich Region of Mitochondrial DNA in the Drosophila subobscura
            Subgroup
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 973)
  AUTHORS   Gonzalez,A.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-FEB-1999) Gonzalez A.M., Department of Genetics,
            Faculty of Biology, University of La Laguna, Tenerife, E 38271,
            SPAIN
FEATURES             Location/Qualifiers
     source          1. .973
                     /organism="Drosophila subobscura"
                     /organelle="mitochondrion"
                     /isolate "Portugal Lisboa subl"
                     /db_xref="taxon:7241"
                     /dev_stage="adult"
     tRNA            <1. .10
                     /gene "tRNA Ile"
                     /evidence=not_experimental
                     /product="transfer RNA Ile"
     gene            1. .10
                     /gene="tRNA-Ile"
     misc_feature    11. .950
                     /note="A+T-rich region"
                     /evidence=not_experimental
     rRNA            951. .>973
                     /note="partial sequence"
                     /evidence=not_experimental
                     /product="small ribosomal RNA"
BASE COUNT      413 a     29 c     40 g    491 t
ORIGIN


  Query Match             4.1%;  Score 42.8;  DB 33,  Length 973;
  Best Local Similarity   58.6%;  Pred. No. 3.7;
  Matches   89;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy      21 ttattttgccgcatttgctaaccgccctattctttttaaaatttaacgattaagaacgc 80
           |||| |  ||   ||| | | |    | || | |||||| | ||||| ||  | |
Db     810 TTATGTCACCATTTTTGGATTGCGAAATTTTTTATTTAAATTGTTAACTATATAAAGATA 869
```

```
Qy      81 aatttcaatttgcgttctacacaaattaacaattcgtccaaaaagaggtatttaagccgaag 140
             ||| ||  ||    | || | | |||| | |||  || | ||  ||||  |||   ||
Db     870 TATTAAACATTTATATATATATATATTATCTATTAATCTAGAACTACTATATAAAAAAAA 429

Qy     141 atttgccatttttttaatccatttttaaaaaaatacatctgcttta 186
           |         | ||  |  || | || |||| |||   |  ||||
Db     430 AAAAAAATAATAATTTTTTAATATATATAAAATTACTATTAATTTA 475


RESULT   6
AF165926
LOCUS       AF165926   163797 bp    DNA             PRI       02-SEP-1999
DEFINITION  Homo sapiens chromosome 5p13 BAC clone djn085o06 containing NUP155
            gene, complete sequence.
ACCESSION   AF165926
VERSION     AF165926.1  GI:5817858
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 163797)
  AUTHORS   Zhang,X., Yu,J., Hu,S., Dong,W., Gregersen,N., Bolund,L. and
            Yang,H.
  TITLE     Human BAC clone djn085o06 containing human NUP155 gene from
            chromosome 5p13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 163797)
  AUTHORS   Zhang,X., Yu,J., Hu,S., Dong,W., Gregersen,N., Bolund,L. and
            Yang,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUL-1999) Human Genome Center, Institute of Genetics,
            Chinese Academy of Sciences, Datun Road, Beijing 100101, P. R.
            China
FEATURES             Location/Qualifiers
     source          1. .163797
                     /organism "Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5p13"
                     /clone "BAC clone djn085o06"
     mRNA            join(75005. .75122,75385. .75481,81218. .81288,84721. .84813,
                     86189. .86355,87280. .87385,88300. .88373,88949. .89040,
                     94899. .94996,96303. .96455,99625. .99725,103910. .104080,
                     105744. .105854,107407. .107501,108261. .108349,
                     109119. .109181,109763. .109910,111572. .111638,
                     113432. .113547,119457. .119554,123164. .123294,
                     126747. .126938,128223. .128361,130053. .130188,
                     132274. .132427,132641. .132745,134068. .134222,
                     134574. .134703,135931. .136044,137913. .138033,
                     138503. .138613,143031. .143167,144180. .144286,
                     145127. .>145265)
                     /product="NUP155"
     CDS             join(75005. .75122,75385. .75481,81218. .81288,84721. .84813,
                     86189. .86355,87280. .87385,88300. .88373,88949. .89040,
                     94899. .94996,96303. .96455,99625. .99725,103910. .104080,
                     105744. .105854,107407. .107501,108261. .108349,
                     109119. .109181,109763. .109910,111572. .111638,
                     113432. .113547,119457. .119554,123164. .123294,
                     126747. .126938,128223. .128361,130053. .130188,
                     132274. .132427,132641. .132745,134068. .134222,
                     134574. .134703,135931. .136044,137913. .138033,
                     138503. .138613,143031. .143167,144180. .144286,
                     145127. .145265)
                     /codon_start=1
                     /product="NUP155"
                     /protein_id="AAD52966.1"
                     /db_xref="GI:5817859"
                     /translation="MSDMDYPLQGPGLLSVPNLPEISSIRRVPLPPELVEQFGHMQCN
                     CMMGVFPPISRAWLEIIDSDIEMWNYEDGGDLAYFDGLSETILAVGLVKPKACIFQPHV
                     RHLLVLATPVDIVILGLSYANLQTGSGVLNDSISGGMQLIPDPLYSLPTDNTYLLTIT
                     STDNGRIFLAGKDGCLYFVAYQAEAGWFSQRCRKINHSKSSISFLVPSLLQFTFSEDD
```

PTLQIAIDNSRNILYTRSEKGVIQVYDLGQDGQCMSRVASVSQNAIVSAAGNIARTID
RSVFKPIVQIAVIENSESLDCQLLAVTHAGVRLYFSTCPFRQPLARPNTLTLVHVRLP
PGFSASSTVEKPSKVHPALYSKGTLLMAASENEDNDILWCVNHDTFPFQKPMMETQMT
AGVDGHSWALSAIDELKVDKIITPLNKDHIPITDSPVVVQQHMLPPKKFVLLSAQGSL
MFHKLRFVDQLRHLIVSNVGGDGEEIERFFKLHQEDQACAICLLLACSTAACDFEVSA
WAIRAFFRYCGEAQMRFPTTLPPPSNVGPILGSPVYSSSPVPSGSPYPNPSFLGTPSH
GIQPPAMSTPVCALGNPATQATNMSCVTGPEIVYSGKHNGICIYFSKIMCNIWDASLV
VERIFKSGNREIIAIESSVPCQLLESVLQELKGLQEFLDRNSQFAGGPLGNPNTTAKV
QQRLIGFMRPENGNPQQMQQELQRKFHEAQLSEKISLQAIQQLVRKSYQALALWKLLC
EHQFTIIVAELQKEIQEQLKITTFKDLVIRDKELTGALIASLINCYIRDNAAVDGISL
HLQDICPLLYSIDDAIGSKANELLQRSPQVQNKTEKERMLPESLKEYQKISNQVDLSN
VCAQYRQVRFYEGVVELSLTAAEKKDPQCLGLHFYKHGEPEEDIVGLQAFQERLNSYK
CITDTLQELVNQSKAAPQSPSVPKKPGPPVILSSDPNMLSNEEACHHFEQMLKISQRSK
DELFSTALYNWLIQVDLADKLIQVASPFLEPHLVRMAKVQQNRVRYMDLLWRYYEKNR
SFSNAARVLSRLADMHSTEISLQQRLEYIARAILSAKSSTAISSIAADGEFLHELEEK
MEVARIQLQIQFTLQRQYSHHSSVQDAVSQLDSELMDITKCYGEFADPFKLAECKLAI
IHCAGYSDPILVQTLWQDIIEKELSDSVTLSSSDRMHALSCKIVLLGKIYAGTPREFP
LDFIVQFLEQQVCTLNWDVGFVIQTMNEICVPLPRLLEVYQQLFKSRDPFWNRMKKPL
HLLDCIHVLLIRYVENPSQVLNCERRRFTNLCLDAVCGYLVELQSMSSSVAVQAITGN
FKSLQAKLERLH"

BASE COUNT 47095 a 35017 c 33525 g 48139 t 21 others
ORIGIN

Query Match 4.1%; Score 42.6; DB 39; Length 163797;
Best Local Similarity 61.1%; Pred. No. 3.4;
Matches 69; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Qy 88 ttttqcgttctacacaaattaacaattcgtccaaaaaqaggtatttaaqcqaaqatttqqc 147
|||| | |||||| ||| |||| || || | ||| || | | | | |
Db 36786 TTTTAATCTGTACACATTTTACCAAATTATCAAATTGTCATATATACCCCAAAAATATAT 36845

Qy 148 attttttaatccattttaaaaaaatacatctqctttaacccacctttqcca 200
|| || || ||| |||||| | ||| || | ||| ||| ||| | | |||
Db 36846 ATATTATTTATCAATTTTTTAGAAACTAAAACTGTCATAATCCAGCAATCCCA 36898

RESULT 7
AB026647
LOCUS AB026647 81542 bp DNA PLN 07-MAY-1999
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 3, P1 clone: MJL12, complete sequence.
ACCESSION AB026647
VERSION AB026647.1 GI:4757403
KEYWORDS HTG.
SOURCE Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1 clone:MJL12.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons, core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (sites)
AUTHORS Nakamura,K.
TITLE Structural Analysis of Arabidopsis thaliana Chromosome 3. II
JOURNAL Unpublished (1999)
REFERENCE 2 (bases 1 to 81542)
AUTHORS Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (28-APR-1999) to the DDBJ/EMBL/GenBank databases. Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of Gene Structure 2; Yana 1532-3, Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp, Tel:81-438 52-3935(ex.7443), Fax:81-438-52-3934)
FEATURES Location/Qualifiers
source 1..81542
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="3"
/clone="MJL12"
/clone_lib="Mitsui P1"
BASE COUNT 25859 a 14006 c 14834 g 26843 t
ORIGIN

Query Match 4.0%; Score 41.8; DB 7; Le
Best Local Similarity 56.0%; Pred. No. 5.4;
Matches 82; Conservative 0; Mismatches 67; I

Qy 52 ctttttaaaaatttttaacqattaaqaacqcaattcaatttqacqttcc
|| |||| |||| || |||||| ||| |||| ||
Db 77374 CTGCTTAACATTTCAAAATTTAAGAAACTAATCTCTTTTTTTTTT

Qy 112 attcgtccaaaaqaqgtatttaaqcqaaqatttqqcatttttttaa
|| | || || | |||| || || | ||| || || ||
Db 77434 ATCTCTTAAAGAGTAAATTATAACAAAATTTATGTCATATTATTGC

Qy 172 aatacatctqctttaacccacctttqcca 200
| || || | ||| || | || ||
Db 77494 ATTATATATTTTTTCCCAAA GTTTACA 77522

RESULT 8
AC007396/c
LOCUS AC007396 97208 bp DNA PLN
DEFINITION Genomic sequence for Arabidopsis thaliana B 1, complete sequence.
ACCESSION AC007396
VERSION AC007396.5 GI:5019263
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Emb euphyllophytes; Spermatophyta; Magnoliophyt eudicots; Rosidae; eurosids II; Brassicales Arabidopsis.
REFERENCE 1 (bases 1 to 97208)
AUTHORS Chao,Q., Shin,P., Brooks S., Buehler,E., Walker,M., Brooks,S., Altafi,H., Araujo,R. Gonzalez,A., Hansen,N.F., Huizar,L., Kremet Li,J., Liu,S., Luros,S., Rowley,D., Schwar Vysotskaia,V., Yu,G., Davis,R.W., Federspi Ecker,J.R.
TITLE Genomic sequence for Arabidopsis thaliana [
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 97208)
AUTHORS Ecker,J.R.
TITLE Direct Submission
JOURNAL Submitted (26-APR-1999) Arabidopsis thalia Department of Biology, University of Penns Hamilton Walk, Philadelphia, Pennsylvania
REFERENCE 3 (bases 1 to 97208)
AUTHORS Ecker,J.R.
TITLE Direct Submission
JOURNAL Submitted (08-JUN-1999) Arabidopsis thalia Department of Biology, University of Penns Hamilton Walk, Philadelphia, Pennsylvania
REFERENCE 4 (bases 1 to 97208)
AUTHORS Chao,Q., Brooks,S., Buehler,E., Johnson-B Shinn,P., Altafi,H., Bei,B., Chin,C., Chi Conway,A., Gonzalez,A., Hansen,N., Howing Lee,J., Jones,C., Li,J., Liu,A., Liu,J., L Nguyen,M., Palm,C., Pham,P., Saikang,F., S Shavers,A., Taylor,B., Vysberg,M., Yu,G Federspiel,N., Theologis,A. and Ecker,J.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-2000) Arabidopsis thali Department of Biology, University of Penn Hamilton Walk, Philadelphia, PA 19104-601
COMMENT On Jun 8, 1999 this sequence version repl
FEATURES Location/Qualifiers
source 1..97208
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"

```
                /chromosome="1"
     CDS        complement(join(449. .1125,1197. .1487,2277. .2673))
                /note "putative tubulin beta-1 chain sp|P12411 similar to
                EST qb|W43629"
                /codon_start=1
                /evidence not_experimental
                /product="T4012.1"
                /protein_id "AAF26774.1"
                /db_xref "GI:6721120"
                /translation="MPEIIHVQAGQCGNQIGSKFWEVICDEHGIDPTGRYNGDSALQ
                LERINVYYNEASGGKYVPRAVLMDLEPGTMDSIRSGPFGQIFRPDNFVFGQSGAGNNW
                AKGHYTEGAELIDAVLDVVRKEAENCDCLQGFQVCHSLGGGTGSGMGTLLISKIREEY
                PDRMMLTFSVFPSPKVSDTVVEPYNATLSVHQLVENADECMVLDNEALYDICFRTLKL
                STPSCGYHQNLNVPGIINELISAIMSCVTCSLRFPGQLNSDLRKLAVNLIPFPRLHFF
                MVGFAPLTSRGSQQYRSLTVPELTQQMWDAKNMMCAADPRHGRYLTASAMFRGRMSTE
                EVDEQIINVQNKNSSYFVEWIPNNVKSSVCDIPPTGLKMASTFVGNSTSIQEMFKRVS
                EQFTAMFRRKAFLHWYTGEGMDEMEFTEAESNMNDLVSEYQQYQDATADIEGEDEEAE
                EQVYES"
     CDS        complement(join(4051. .4260,4340. .4418,4501. .4866,
                4970. .5330,5428. .5530,5743. .6013,6087. .6196,6322. .6459))
                /note="putative pectinesterase like protein
                emb|CAA18104.1"
                /codon_start=1
                /evidence not_experimental
                /product="T4012.2"
                /protein_id="AAF26773.1"
                /db_xref="GI:6721119"
                /translation="MRHVFVEVLVLISLVILELSYAFAPISSYQWVVSYSQREILGGN
                KQVIVINDMFPCPILNATANDIIVVNIPNNLPEPFLMTWNGLQLRKNSWQDGVRGTNC
                PILPGTNWTYRFQVKDQIGSYFYFPTLLLQKAAGGYGAIRIVPPFLVPVPFPKPDDEY
                DILIGDWFYLDHIVMRASLDAGHSLPNPDGILPNGRGPEPTPEAPPGKTYRLRIRSNV
                GLKICLNFRIQNHNMLIVETEGTYVQKEVYSSLDIHVGQSYSILVTAKTPGDYWIVS
                YIFATARFIDSYLSGIALIRYPGSPLDPVGQGPLAPALQDFGSSVRQALSIPMDLNVG
                AARSNPQGSYHYGRINVTRTIILRNDVMIESCKLEYTINGVSFVYPDFTPIFKVLDGI
                NDTIIPGMFPVYPSNKTPTLGTSVVDIHYKIDFIHIVFQNPLFGLPSVHIDGYNPPVVG
                YGFGAWSPKKASNVGVFAVESFSVQVTIYSWLAILIAMGNQGMWNVEQRALAWS
                QFLYMPVVGGPSFPGPSTIDVVPPNPIFGNVIWQPVF"
     CDS        join(7960. .8014,8096. .8798,9284. .9518)
                /note="thaumatin-like protein qb|AAD02499; similar to ESTs
                qb|AI994380.1, qb|AI994188.1 qb|AA650797, qb|T46212"
                /codon_start=1
                /evidence not_experimental
                /product "T4012.3"
                /protein_id="AAF26752.1"
                /db_xref "GI:6721098"
                /translation "MAIPLPLIFIFSHIFVSGVPSTSFIMVNKCPYTVWPGILSNAV
                VPPLPTTGFVLQKGFERTISAPTSWGGRFWGRTQCSTDTDGKFPCLTGDGSGTLECS
                GSGAIPPATLAEFTIDGSNGLDFYDVSLVDGYNVPMLVAPQGGSGLNGSSTGCVVDLN
                CSGPSELKFVSLVGKGKQSMGCKSACFAFKPDERGSCACVTPDPDPKSYSLMFKSA
                CPRAYSYAYDDQSSTFTCAESENYVITFCPTPNTSQKSSQDQSPDPKPTTPTGTSSFT
                PAGDSSTTWSPVDTSMIYFGALDQNKGSPSTCHLSLCGITVTLALAFCPMWRLF"
     CDS        10818. .11195
                /note "putative small hydrophobic protein qb|AAD02500"
                /codon_start=1
                /evidence not_experimental
                /product="T4012.4"
                /protein_id="AAF26753.1"
                /db_xref "GI:6721099"
                /translation="MCLMALSDAVLGNLATIYVAVTIAIKAYGLITGRSFSAGFVVVV
                SITAVGVLLAVTLAWDVSRKAAEAVSKYNEVGGRFDLSHRRHHHDSGAICKQGICWHC
                VAVKSPASQVRFRLPQHTPYGAF"
     CDS        complement(join(12702. .13054,13134. .15717))
                /note="putative CIVI receptor kinase gi|2260756, similar
                to EST qb|AI993519.1"
                /codon_start=1
                /evidence not_experimental
                /product "T4012.5"
                /protein_id="AAF26772.1"
                /db_xref "GI:6721118"
                /translation="MRLLKTHLLFLHILYLFFSPCFAYTDMEVLLNLKSSMIGPKGHGL
                HDWIHSSSPDAHCSFSGVSCDDDARVISLNVSFTPLFGTISPEIGMLTHLVNLTLAAN
                NFTGELPLEMKSLISLKVLNISNNGNLTGTFPGELKAMVHLEVLDLYNNNFNGKLPP
                EMSELEFLFYLSFQNNFFSGEIPESYGDIQSLEYLSLNGNAGLSGKSDAFLSNLKNLEE
```

```
                MYICYYNSYTGGVPPEFGGLTKLETLDMASCTLIGEIPTSLSNLKHLHTLFLHINNLI
                GHIPPELGLVSLKGLDLSINQLTGHIPQSFINLSNLTLLNLFSNNLYGQIPEAICEL
                PKLEVFFVWENNFTLQLPANLGRNGNLIFLDVSDNHLTGLIPKSLCRCEKLEMLILSN
                NFFFGPIPEELGKCKSLTKIRIVKNLLNGTVPAGLFNLPLVTIIELTDNFFSGELPVT
                MSGDVLGQIYLSNNWFSGEIPPAIGNFPNLQTLFLDRNRFFGNIPREIFELKHLSKIN
                TSANNITGGIPSGISKCSKLSVIDLSFNEISGLIPKGINNVRNISTLNLSGNQLTGSI
                PTGIGNMTSLFTLDLSFNDLSGRVPLGGQFLVFNETSFAGNTYICLPHRVSCPTRPGQ
                TSDHNRIAIFSPSFEVIIVIAAIGFLIISVAIFQMNKKENQFSLAWKLIAFQKLDFK
                SEDVIECLFEENIIGKGAMGIVYFGSMPNNVEVAIKELVGRGTGRSDHGFTAEIQTLG
                RIRHRHIVRLLGYVANKETNLLLYEYMPNGSLGELLHGKKGGHLQWETRHRVAVEAAK
                GLCYLHHDCSPLILHRDVESNNILLDSDFEAHVADFGLAKFLVDGAASECMQSIAGSY
                GYIAPEYAYTLKVDEKSDVYSFGVVLLELITGKKPVGEFGEGVDIVRWVRNTEEEITQ
                PSDAATVVAIVDPRLTGYPLTSVIHVFKIAMMCVEEEAAARPTMREVVHMLINPPASV
                ANLIAF"
     mRNA       join(2476. .2593,2641. .2825)
                /note="similar to cysteine rich antifungal protein 1
                precursor sp|P30224, similar to ESTs emb|Z27258,
                emb|Z29957, qb|AA713157"
                /codon_start=1
                /evidence not_experimental
                /product "T4012.6"
                /protein_id="AAF26754.1"
                /db_xref="GI:6721100"
                /translation "MAFPAIVVLFFAALVFFAASTAQMVVEAQEICLRFSGIWSGVC
                GNSNACKNQCINEIKARHGSCNYVFPAHKCICYFPC"
     CDS        join(25035. .25133,25466. .25589,25677. .25786,25899. .25962,
                26045. .26109,26188. .26253,26350. .26448)
                /note="similar to ATGP3 gi|4097565"
                /codon_start=1
                /evidence not_experimental
                /product "T4012.7"
                /protein_id "AAF26755.1"
                /db_xref "GI:6721101"
                /translation "MSASFFIFCVIVCLQAVSELIMLISYISNTIFIENKLYDLVTEY
                QDYVPTVFDNFSANVVVPGNIVNIGIWFTAQGFDYNFIFRLSYPGADVFILAFSLISK
                ASYENVAFAWTELLFHYALCVFLVCLAIIRPLFFLMRDNRCVFL"
     CDS        complement(join(27124. .27233,27435. .27518,27694. .27796,
                27882. .27989,28095. .28175,28432. .28412,28718. .28804,
                28913. .29002,29239. .29297,29507. .29572,29649. .29888,
                30088. .30204,30720. .30812,30862. .30936,31335. .31457,
                31550. .31639,31697. .31780,31963. .32050,32128. .32313,
                32467. .32771,32850. .32945))
                /note="similar to mouse MEM3 gi|2145519"
                /codon_start=1
                /evidence not_experimental
                /product "T4012.8"
                /protein_id="AAF26771.1"
                /db_xref "GI:6721117"
                /translation="MKTLAGVEDEDKWLAEGTAGTQHNAFFMHKALDANNLREVLKYS
                ALMLSELPTSKLSPQKYYDLGFDPQFTLNLSIHACVDQFLIIFPSNLDMRAFDQLRQ
                LEIFFKDFSPHGLPVVDLYFLVQHAGNTLPRMYLICTVGSVYIKSKQAPSKDVLKDLV
                EMGFGVQHFIRGIFIREYIAQVSPDKLEEIGSPTEGLANIVMDAVEPVLQNFTEMNKL
                WVRIQHQGFGTVKEKQEKERNELRDLVISIFLLSLVGKNLHVLGQIEGVDLRMYKETV
                LPRVLEQVVNCKDKLAQYYLMECIIQVEPDEYHIQILELLAACTQLMPTVDTKIVLI
                QLMDRLSNYAASSPQLDVPLFLTSFQVLEFLQVEAFARLSNAIGKVIDTQLEMPIVG
                AMTLFVSLLTFTLRVHPDRLDYVDQVLGACVVKLSSVPKLEDARAMKQVVALLSAPLE
                KYSDIVTALTLSNYPRVMDHLDDGTNKVMAMLIQSIMKTDSCISTADKVEVLFELIK
                GLIKDLDETNAEELQDEEDPQEEQNGVAELIHMLENEEFEEMLKVKHIKFTKATCSLNL
                NMPFIRCTIICVVPPHLMTGCPPPTIPFTVPPLVFSAVPLVPQLESQCCIIACEDSATP
                RKIFQILNQTIEVLTSVPCPELALRLYLQCAEVIYRAASDCDLEPVAYEFFTQAFMLY
                EEEIADSKAQVTAIHLIVGTLQRINVFGVENRTFLTHKATGYSARLLEKPDQCRAVYA
                CSHLFWVDGFEGIFDGFFVLDILEPAIFIANAAQLMASALFCSSGFVILFVEILNKYI
                YFFEKGNPHITPSDIQSLIELINNEMQSDNGNTTIHSDPFFTSTLRYIKFIKQKGGLM
                GEKYDPIKL"
     CDS        complement(join(33326. .33819,33915. .34271,34383. .34449))
                /note="hypothetical protein"
                /codon_start=1
                /evidence not_experimental
                /product="T4012.9"
                /protein_id="AAF26770.1"
                /db_xref="GI:6721116"
                /translation="MCSLVQVLVSCIVFHSVNVEGFLYSNGLKLEETILLSK
```

KAHRIERK*EEEEEKKKKK-RREESHKHS*AAIDNimK:;E:SRKVSITSDSSLDKSC
LIDELEEPQKHGVI SGSSNSKEEFEDESPIAVESLIKAAIEVACKIPLEISMVEKKEE
EEVPILPREAVVCSTTVAKSISHQDVITSSISSSKTSELEKNLPSTSIAAIDETKKRK

Query Match 4.0%; Score 41.6; DB 49; Length 97208;
Best Local Similarity 54.6%; Pred. No. 6;
Matches 83; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

Qy 53 ttttttaaaatttaacqattaaqaaacqcaattcaatttqcqttctacacaaattaacaa 112
||||| || | || ||||| || | || || | || | |||| || ||
Db 53893 TTTTTTAATCTAAAAATATTAAAAATTAACTTTAAATAAAAATTGCAAGAAATAAAAAA 53834

Qy 113 ttcqtccaaaaqaqqtatttaaqccaaqatttqqcatttttttaatccatttttaaaaaa 172
| ||| ||| | | ||||| || | | || | || | | | | |||| ||
Db 53833 TCCGTGAGAAAAAGCCTTTAATTAAAATTATAAAATATATTAATTTCTTAACTAAACAA 53774

Qy 173 atacatctqctttaaccacctttqccaaaqq 204
| || ||| | | | | | ||| |
Db 53773 GAAAATGGGCTGACAATTAACCTCCCAATCCC 53742

RESULT 9
AC022522/c
LOCUS AC022522 88643 bp DNA PLN 04-MAR-2000
DEFINITION f28K15, complete sequence.
ACCESSION AC022522
VERSION AC022522.2 GI:6978637
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 88643)
AUTHORS Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F., Altafi,H., Nguyen,M., Lam,B., Southwick,A., Miranda,M., Brooks,S., Buehler,E., Chao,Q., Chin,C., Chiou,J., Choi,E., Gonzalez,A., Howng,B., Johnson-Hopson,C., Khan,S., Kim,C., Koo,T., Lee,J.M., Lenz,C., Liu,A., Liu,S., Mukharsky,N., Pham,P., Sakano,H., Shinn,P., Toriumi,M., Vaysberg,M., Yu,G., Ecker,J., Theologis,A and Davis,R.W.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 88643)
AUTHORS Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F., Altafi,H., Nguyen,M., Lam,B., Southwick,A., Bei,Q., Buehler,E., Chin,C., Chiou,J., Choi,E., Dunn,P., Gonzalez,A., Howng,B., Kim,C., Koo,T., Lee,J.M., Lenz,C., Li,J., Liu,A., Liu,F., Liu,S., Mukharsky,N., Pham,P., Sakano,H., Schwartz,J., Shinn,P., Thaveri,A., Toriumi,M., Vaysberg,M., Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
TITLE Direct Submission
JOURNAL Submitted (05-FEB-2000) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA
REFERENCE 3 (bases 1 to 88643)
AUTHORS Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F., Altafi,H., Nguyen,M., Lam,B., Southwick,A., Bei,Q., Buehler,E., Chin,C., Chiou,J., Choi,E., Dunn,P., Gonzalez,A., Howng,B., Kim,C., Koo,T., Lee,J.M., Lenz,C., Li,J., Liu,A., Liu,K., Liu,S., Mukharsky,N., Pham,P., Sakano,H., Schwartz,J., Shinn,P., Thaveri,A., Toriumi,M., Vaysberg,M., Walker,M., Yu,G., Ecker,J., Theologis,A and Davis,R.W.
TITLE Direct Submission
JOURNAL Submitted (04-MAR-2000) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA
COMMENT On Feb 14, 2000 this sequence version replaced gi:6910565.
Bases 57,901-88,643 of clone f28K15 overlap with bases 1-30,740 of IGF BAC clone F12F1 (AC002131).
FEATURES Location/Qualifiers
source 1..88643
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="1"
/map="bacf N°1"
/clone="f28K15"
BASE COUNT 29756 a 15417 c 14998 g 28672 t
ORIGIN

Query Match 3.9%; Score 40.8; DB 49; L
Best Local Similarity 51.7%; Pred. No. 9.4;
Matches 93; Conservative 0; Mismatches 87; G

Qy 16 tatatttattttqccqcattttqctaacqccctattcttttttaaa 65
||||| ||| || | | |||| | | || | ||| ||| ||
Db 72458 TATATATATGTTAACACGTTTTTATGAACGGCGATTTTTTGAAATA 72409

Qy 76 aacqcaattcaatttqcattctatcacaaattaacaattcqtccq 121
| | ||| | | | || || | | |
Db 72398 TATGACATTAATAAAAGACTGTTTTATATTGTTACCGTTTTACAA 72353

Qy 136 cqaaqatttqqcatttttttaatccattttaaaaaaatacatct 180
|| || ||| | | || ||| ||||| |
Db 72298 TTTACGTTCCTCATTATATTAACAATCAACTAATAAAATAGTTC 72254

RESULT 10
F12F1/c
LOCUS F12F1 12486 bp DNA PLN
DEFINITION Arabidopsis thaliana chromosome 1 BAC F12F1 sequence.
ACCESSION AC002131
VERSION AC002131.1 GI:2564872
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 12486)
AUTHORS Vysotskaia,V.S., Osborne,B.I., Schwartz,J., Kwan,A., Oji,Y., Liu,S., Buehler,E., Conway,A., Dewar,K., Fedorov,I., Kim,C., Kurtz,D., Palm,C., Sun,H., Davis,R.W., Ecker,J.R., Federspiel,N.
TITLE Arabidopsis thaliana chromosome 1 BAC F12F1 sequence
JOURNAL Unpublished (1998)
REFERENCE 2 (bases 1 to 12486)
AUTHORS Theologis,A.
TITLE Direct Submission
JOURNAL Submitted (20 MAY 1997) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA
REFERENCE 3 (bases 1 to 12486)
AUTHORS Theologis,A.
TITLE Direct Submission
JOURNAL Submitted (04 NOV-1997) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA
REFERENCE 4 (bases 1 to 12486)
AUTHORS Theologis,A.
TITLE Direct Submission
JOURNAL Submitted (27 MAY 1998) Plant Gene Expression Center, 800 Buchanan St., Albany, CA 94710, USA
REFERENCE 5 (bases 1 to 12486)
AUTHORS Theologis,A.
TITLE Direct Submission
JOURNAL Submitted (28 MAY 1998) Plant Gene Expression Center, 800 Buchanan St., Albany, CA 94710, USA
COMMENT On Nov 4, 1997 this sequence version repl
.
FEATURES Location/Qualifiers
source 1..12486
/organism="Arabidopsis thaliana"
/cultivar="Columbia"

```
                     /db_xref="taxon:3702"
                     /chromosome="1"
                     /clone="F12F1"
     gene            complement(5335..9284)
                     /gene="F12F1.1"
     CDS             complement(join(5335..6199,7924..8498,8673..8712,
                     8807..8984,9170..9284))
                     /gene="F12F1.1"
                     /note="similar to ubiquitin-activating enzyme E1
                     qb|L10357 from A. thaliana. ESTs qb|H37523 and qb|H37524,
                     qb|T46117, qb|T46068, qb|T75688, qb|H29817, qb|H29862,
                     qb|Z34634 and qb|Z34258 come from this gene."
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAC17604.1"
                     /db_xref="GI:3157921"
                     /translation="MSIPETKSPDDIIIDAWDPQGPADPSFTGCWASAAMILGIEAVE
                     RLFILGIGVNLVTYLPGIMELGNATAANIVTNFLGISFMLCLLGCFIADTFLGRYLTI
                     AIFAAIQAIGVSLLTLSTIIPGLRPPPCNPTTSSHCPQASGIQLTVLYLALYLIALGT
                     GGVKASVSGFGSDQFDETDPKERSKMTYFFNRFFFCINVGSLLAVTVLVYVQDDVGRK
                     WGYGICAFAIVLALSVFLAGTNPYFFKKLIGSPMTQVAAVIVAAWRNRKLELPADPSY
                     LYDVDDIIAAEGSMKGKQKLPHTFQFRSIDKAAIRDQEAGVTSNVFNKWTLSTLTDVE
                     EVKQIVRMLPIWATCILFWTVHAQLTTLSVAQSFTLDPSIGSFEIPPASMAVFYVGGI
                     LLITAVYDRVAIRLCKKLFNYPHGLRPLQRIGLGLFFGSMAMAVAALVELKRLRTAHA
                     HGPIVKTLPIGFYLLIPQYLIVGTGEAITYIGQLDFFLRECPKGMKGMGTGLFLSTLA
                     LGFFFSSVLVTIVEKFTGKAMPWIADDLNKGKLYNFYWLVAVLVALNFLIFLVFSKWY
                     VYKEKKLAEVGIELDDEPSIPMGH"
     gene            complement(18780..19127)
                     /gene="F12F1.2"
     CDS             complement(18780..19127)
                     /gene="F12F1.2"
                     /note="Contains similarity to proline-rich protein,
                     qb|S68113 from Brassica napus."
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAC17605.1"
                     /db_xref="GI:3157922"
                     /translation="MFPAHVSQCNTCFPTPTPIPNLNPISNPTTPGCSRDAIKLGVCAK
                     ILDVAVGTVIGNPSDTLCCSVLQGLVDLDAAVCLCTTIKANILGINIDLPISLSLLIN
                     TCGKKLPSDCICA"
     gene            24036..24449
                     /gene="F12F1.3"
     CDS             24036..24449
                     /gene="F12F1.3"
                     /note="Contains homology to extensin like protein
                     qb|D83227 from Populus nigra. ESTs qb|H76425, qb|T13883,
                     qb|T45348, qb|H37743, qb|AA042634, qb|Z26960 and qb|Z25951
                     come from this gene. There is a similar ORF on the
                     opposite strand."
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAC17607.1"
                     /db_xref="GI:3157924"
                     /translation="MASPSIALPLAINLIFTFTLCAVSSTIPSQQCDPSPKIPPSLPP
                     IPSPSSGSSKCPKDFLKLGVCANVLNCLLDNLPLCKPPVSPPCSLLQCLADVEAAVCLC
                     TALKANILGINLNLPISLSLLLNVCSKQLPPGFQC"
     gene            complement(29737..30836)
                     /gene="F12F1.4"
     CDS             complement(join(29737..30138,30322..30836))
                     /gene="F12F1.4"
                     /note="ESTs qb|H12979, qb|AA042648, qb|H37240, qb|T76189,
                     qb|H36952, qb|T22022 and qb|Z42279 come from this gene. A
                     similar ORF is present on the opposite strand."
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAC17608.1"
                     /db_xref="GI:3157925"
                     /translation="MATVPVFEQVTPVAVPNVPVPTKPVPPTVVPTPVTAQDPPSVPAV
                     TEQKSEAPIVETNEEVVVEEAEKKDEETEKPTPPPTPTPVITPTPVVEEEEEKAEEV
                     TETPAVVEEEKKTEVVPEKQTEVAAAEDVAVEKAEE"
     gene            34519..35712
                     /gene="F12F1.5"
     CDS             join(34519..34701,34793..34859,35037..35261,35354..35478,
                     35641..35712)
                     /gene="F12F1.5"
                     /note="Contains similarity to GDP-dissociation inhibitor
                     qb|L07918 from Mus musculus."
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAC17610.1"
                     /db_xref="GI:3157927"
                     /translation="MVINEELMFASSSIDPKQTSFFNSSSSMFTTSSLEEEEEEE
                     LELGFMIALPEQLEEEEDEELEEWEEQMLDFVLEEVGETTDFLVEDLTLTIRSHIK
                     EDMVLTIFDGSFAGKIWFTLEEGSFYTLTTTDTVTNNIVSSLQYNIVWKIIKVY
                     SRKLMGTTPSQAEPYTHVMFEETAPSGLLVRGVSVASKTVDDDNQCYLENNYIFDI
                     RKNWI."
     gene            36591..37248
                     /gene="F12F1.6"
     CDS             join(36591..36897,37076..37248)
                     /gene="F12F1.6"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="F12F1.6"
                     /protein_id="AAC17603.1"
                     /db_xref="GI:3157920"
                     /translation="MFKTYFIQSLFSSLITLIIILLISRAAAVGSGGVCRHPPSQNSC
                     KPCMAKQMKYPVCVRVLKCMAKCMGGCVTQRKYTTTCRCGIAAKPSLLECKRCVSR
                     PPIYSHISKRYITLPFPPGQYTHTHTHILMNLTPDYDTSILFRSIYSHTHSNKIELIN"
     gene            complement(37644..38291)
                     /gene="F12F1.7"
     CDS             complement(37644..38291)
                     /gene="F12F1.7"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="F12F1.7"
                     /protein_id="AAC17606.1"
                     /db_xref="GI:3157923"
                     /translation="MKRSKRFSSSTTITIVIHIFYNDHIIPPAIKEIPIEIPLPAIKA
                     NVKTNATAAAAKIQSGYRSYRIRNLYKKISSINREANPVQSIIQRQETVDAIRSDEKF
                     KLRMNGTLMALLKLDSVFGLDFTTEAAKKVSRKTVGMQEILDSTSELKDEIQWWNY
                     NDLGGVDSGQGGGAWPLYWEEAVEEEMCKERGGEEMERFCAQYLGFRCFQRFLRE"
     gene            complement(38769..41721)
                     /gene="F12F1.8"
     CDS             complement(join(38769..38831,38906..39091,39186..39260,
                     39535..39597,39676..39746,39841..39900,39984..40083,
                     40202..40277,40521..40595,40692..40916,40909..41015,
                     41061..41139,41515..41721))
                     /gene="F12F1.8"
                     /note="Similar to fumarylacetoacetate hydrolase, qb|L41670
                     from Emericella nidulans."
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAC17611.1"
                     /db_xref="GI:3157928"
                     /translation="MALLKSFIDVGSDSHFPIQNLPYGVFKPESNSTPPPAVAIGDLV
                     LDLSAISEAGLFDGLILKDADCFLQPNLNKFLAMGPPAWKEARSTLQRILSFLLFGPK
                     VLVLVCFHAANPILLKGNGVLREESFHQMSSAVPMLVPMVIGDYTDFFASMHHAKNCGL
                     MFRGPFNANPNWFKLPIAYHGRASSIVISGTDIIKRKGQGHPQGNSEPYFGPSKKLD
                     FELEMAAVVGPGNELGKPIDVNNAADHIFGLLLMNDWSARDIQAWEYVPLGPFLGKSF
                     GTTISPWIVTLDALEPFGCQAPKQDPPPLPYLAEKESVNYDISLELAHITVNGCNLRP
                     GDLLGTGTISGPEPDSYGCILELTWNGQKPISLNGETKTPFFDGDQVIFSGVCKGDGY
                     NVGFGTCTGKIVPSPP"
     gene            complement(42231..44465)
                     /gene="F12F1.9"
     CDS             complement(42231..44465)
                     /gene="F12F1.9"
                     /note="Strong similarity to extensin like protein
                     qb|Z34465 from Zea mays."
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAC17609.1"
                     /db_xref="GI:3157926"
                     /translation="MEFPPLRSLFIFTLLISSVCFLQIKAPHHDFSDLGSDFKVDKRL
                     KFENPKLKQAYIAIQSWKKAIFSDPFNFTANWNGSDVCSYNGTYCAPSPSYPKTRVVA
                     GIDLNHADMAGYLASEGLLSDLALFHINSNRFCGEVPLTFNSMKLLTFLDLSNNRFV
                     GKFPKVVLSLPSLKYLDLRYNEFDGKIPSNLFDKDLDAIFLNHNEFRGTIPKNMGNSP
```

```
                    VSALVLADNNLGGCIPGSIGQMGKTLNELILSNDNLTGCLPPQIGNLKKVTVFDITSN
                    RLQGPLPSSVGNMKSLEELHVANNAFTGVIPPSICQLSNLENFTYSSNYFSGRPPICA
                    ASLLADTVVNGIMNCTTGLARQRSDAQCSSLLARPVDCSKGGCYNIFSPPPPTFKMSP
                    EVRTLPPPIYVYSSPPPPPSSKMSPTVRAYSPPPPPSSKMSPSVRAYSPPPPPYSKMS
                    PSVRAYPPPPPSPSPPPPYVYSSPPPPYVYSSPPPPPYVYSSPPPPPYVYSSPPPPY
                    VYSSPPPPYVYSSPPPPPPSPPPPCPESSPPPPVVYYAPVTQSPPPPSPVYYPPVTQS
                    PPPPSPVYYPPVTNSPPPPSPVYYPPVTYSPPPPSPVYYPQVTPSPPPPSPLYYPPVT
                    PSPPPPSPVYYPPVTPSPPPPSPVYYPPVTPSPPPPSPVYYPSETQSPPPPTEYYYSP

  Query Match              3.9%;  Score 40.8;  DB 8;  Length 123386;
  Best Local Similarity   51.7%;  Pred. No. 9.3;
  Matches   93;  Conservative    0;  Mismatches   87;  Indels    0,  Gaps    0;

Qy      16 tatatttatttqccqcatttttqctaaccqcctattcttttaaaaattttaacqattaaq 75
           ||||| ||| ||  | | |||  | | || | ||| |||  || | |||    |||
Db  14558 TATAIATATGTTAACACGTTTTTATGAACGGCGATTTTTTCAAATAATTTTCAAATTTTA 14499

Qy      76 aacqcaattcaatttttqcqttctacacaaattaacaattcqtccaaaaqatqtatttaaq 135
           | |  ||| |     |  |    || |   | ||  ||   | |   ||   | |||
Db  14498 TATGACATTAATAAAAGACTGGGGCATATTGTTACCGTTTTACAACTTGAATAACTTAGC 14439

Qy     136 cqaaqatttqqcatttttttaatccatttttaaaaaaatacatctqctttaacccacctt 195
           || ||   |||| | |||          ||| ||||||  || || |||    ||||||
Db  14438 TTTAGCTTCCTCATTATATTAGCAATCAACTAATAAAATAGGTCGGCATTACTTTACCTT 14379


RESULT  11
AC004153/c
LOCUS       AC004153   153418 bp    DNA             HTG       15-MAR-2000
DEFINITION  Plasmodium falciparum chromosome 12 clone 3D7, *** SEQUENCING IN
            PROGRESS ***, 4 unordered pieces.
ACCESSION   AC004153
VERSION     AC004153.5  GI:7243837
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 153418)
  AUTHORS   Hyman,R.W., Fung,E.L., Qin,F., Rowley,D., Tamaki,T., Kurdi,O.B.,
            Conway,A.B. and Davis,R.W.
  TITLE     Plasmodium falciparum 3D7 chromosome 12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 153418)
  AUTHORS   Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-FEB-1998) Stanford DNA Sequencing and Technology
            Center, Stanford University, 855 California Avenue, Palo Alto, CA
            94304, USA
COMMENT     On Mar 15, 2000 this sequence version replaced gi:6652491.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    42573: contig of 42573 bp in length
            *   42574    42773: gap of unknown length
            *   42774    93373: contig of 50600 bp in length
            *   93374    93573: gap of unknown length
            *   93574   149671: contig of 56098 bp in length
            *  149672   149871: gap of unknown length
            *  149872   153418: contig of 3547 bp in length.
FEATURES             Location/Qualifiers
     source          1..153418
                     /organism="Plasmodium falciparum"
                     /db_xref="taxon:5833"
                     /chromosome="12"
                     /clone="3D7"
BASE COUNT    60163 a   14509 c   15093 g   63052 t     601 others
ORIGIN


  Query Match              3.9%;  Score 40.6;  DB [illegible];
  Best Local Similarity   58.8%;  Pred. No. 10;
  Matches   70;  Conservative    0;  Mismatches   49;

Qy      48 tattcttttaaaattttaacqattaaaaacqcaattcaattttq 93
           | || |||||    |||| | | | ||   |||  ||| ||
Db  98256 TTTTTTTTTTTTTTTTTTTCAATAATTAAAAGAATATAATATTA 98207

Qy     108 aacaattcqtccaaaaqatqtatttaaqcqaaqatttqqcattttt 153
           | | || ||  |||. |  ||| |  .      ||| || ||.
Db  98196 AGGATTTGTAAAAAAATATTTATTTTTGTTTTTTTTTT-ATTTA 98151


RESULT  12
AF153947/c
LOCUS       AF153947      474 bp    DNA             INV       [illegible]
DEFINITION  Euphydryas desfontainii 16S ribosomal RNA [illegible]
            mitochondrial gene for mitochondrial product.
ACCESSION   AF153947
VERSION     AF153947.1  GI:6002850
KEYWORDS    .
SOURCE      Euphydryas desfontainii.
  ORGANISM  Mitochondrion Euphydryas desfontainii
            Eukaryota; Metazoa; Arthropoda; Tracheata;
            Pteryqota; Neoptera; Endopteryqota; Lepidoptera;
            Ditrysia; Papilionoidea; Nymphalidae; Nymphalinae; Euphydryas.
REFERENCE   1  (bases 1 to 474)
  AUTHORS   Zimmermann,M., Wahlberg,N. and Descimon,H.
  TITLE     A phylogeny of Euphydryas s.l. checkerspot butterflies
            (Lepidoptera: Nymphalidae) based on mitochondrial DNA
  JOURNAL   Ann. Entomol. Soc. Am. (1999) In press
REFERENCE   2  (bases 1 to 474)
  AUTHORS   Zimmermann,M., Wahlberg,N. and Descimon,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAY-1999) Division of Population Biology,
            of Ecology and Systematics, University of Helsinki, P.O. Box
            Arkadiankatu 7, FIN-00014, Finland
FEATURES             Location/Qualifiers
     source          1..474
                     /organism="Euphydryas desfontainii"
                     /organelle="mitochondrion"
                     /isolate="[illegible]"
                     /db_xref="taxon:104507"
     rRNA            1..474
                     /product="16S ribosomal RNA"
BASE COUNT      215 a      [illegible] c      95 g     175 t
ORIGIN


  Query Match              [illegible]%;  Score 40.4;  DB [illegible]
  Best Local Similarity   51.1%;  Pred. No. 14;
  Matches   95;  Conservative    [illegible];  Mismatches   [illegible];

Qy       2 atctqtaaattqcaatataattatttqccqcatttttqctaac 46
           |||  |||  ||    | |  |  |  |||    |
Db     379 AICCCTAAGCTAATTTTTTCTTTTAAICATTATTAATCATC [illegible]

Qy      62 ttttaacqattaaaaacqcaattaattttqcqttctacacc [illegible]
           |||||  || || ||    |.| ||||.|    |' ' '
Db     319 TTTTAAAAATAAAAAATTTATTAAAATTTTCTTATCATTTA [illegible]

Qy     122 aagatqtatttaatqaaaaatttqcattttttttaatqcatt [illegible]
                |||||  |  |  |  |  ||  ||  ||
Db     259 TTATTAAATTTAAATTTATATATATATATATATATATATT [illegible]

Qy     182 ctttaa 187
               ||||
Db     199 AATTAA 194


RESULT  13
AF010513
```

```
LOCUS       AF010513      488 bp    DNA             INV       16-JUL-1998
DEFINITION  Plasmodium falciparum microsatellite TA28 sequence.
ACCESSION   AF010513
VERSION     AF010513.1  GI:2253505
KEYWORDS    .
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 488)
  AUTHORS   Su,X.z. and Wellems,T.E.
  TITLE     Toward a high-resolution Plasmodium falciparum linkage map:
            polymorphic markers from hundreds of simple sequence repeats
  JOURNAL   Genomics 33 (3), 430-444 (1996)
  MEDLINE   96299638
REFERENCE   2  (bases 1 to 488)
  AUTHORS   Su,X.z. and Wellems,T.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JUN-1997) NIAID-Laboratory of Parasitic Diseases,
            NIH, Bethesda, MD 20892, USA
FEATURES             Location/Qualifiers
     source          1..488
                     /organism="Plasmodium falciparum"
                     /db_xref="taxon:5833"
                     /chromosome="6"
     repeat_region   1..488
                     /note="microsatellite TA28"
BASE COUNT      182 a     69 c     37 g    200 t
ORIGIN
```

```
  Query Match              3.9%;  Score 40.4;  DB 44;  Length 488;
  Best Local Similarity   50.5%;  Pred. No. 14;
  Matches   98;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy       2 atctgtaaattacatatatttatttttgccgcatttgctaaccgcctattcttttttaaaa 61
           ||||   ||  ||||||||||| ||||     ||||  |   |   | | | | ||
Db     249 ATCTTATAAAAACATATATTTTTTTTTTTTTTTTTTTTTTTTCATTTTTACATCTTGCTC 308

Qy      62 ttttaacgattaagaacgcaatttcaatttgcgttctacacaaattaacaattcgtccaaa 121
           ||| ||   | || |  ||   ||   ||    |  |   || || ||| |  |   |
Db     309 ATTTGACTCATTAGGAACCATAGCATCGTTTATATTGTCTATAAATACCAAATGTTTATA 368

Qy     122 aagaggttatttaagcgaagattgtgcatttttttaatccatttttaaaaaaatacatctg 181
           || | ||    |  || || |    | | |  ||  || | || | | | | || ||
Db     369 AATAAATAAAATAAATAAAATATATATATATATATATATATATATATATACATATATG 428

Qy     182 ctttaacccaccctt 195
           | ||| |||| | |
Db     429 CATTATCCCAACAT 442
```

```
RESULT  14
TETMTRGTG
LOCUS       TETMTRGTG     3439 bp    DNA             INV       07-MAR-1995
DEFINITION  T.pyriformis mitochondrial ribosomal RNA large subunit, Leu- and
            Met tRNA genes.
ACCESSION   M58011
VERSION     M58011.1  GI:343464
KEYWORDS    ribosomal RNA large subunit; transfer RNA-Leu; transfer RNA Met.
SOURCE      T.pyriformis (strain S1) mitochondrial DNA, clone HindIII/Sau3A.
  ORGANISM  Mitochondrion Tetrahymena pyriformis
            Eukaryota; Alveolata; Ciliophora; Oligohymenophorea;
            Hymenostomatida; Tetrahymenina; Tetrahymena.
REFERENCE   1  (bases 1 to 3439)
  AUTHORS   Heinonen,T.Y., Schnare,M.N. and Gray,M.W.
  TITLE     Sequence heterogeneity in the duplicate large subunit ribosomal RNA
            genes of Tetrahymena pyriformis mitochondrial DNA
  JOURNAL   J. Biol. Chem. 265 (36), 22336-22341 (1990)
  MEDLINE   91093076
FEATURES             Location/Qualifiers
     source          1..3439
                     /organism="Tetrahymena pyriformis"
                     /organelle "mitochondrion"
                     /strain="S1"
                     /db_xref="taxon:5908"
                     /clone="HindIII/Sau3A"
     rRNA            670..2984
                     /gene="LSU-beta"
                     /product="large subunit ribosomal RNA"
     gene            670..2984
                     /gene="LSU-beta"
     tRNA            2987..3067
                     /gene="tRNA-Leu"
                     /product="transfer RNA-Leu"
     gene            2987..3067
                     /gene="tRNA-Leu"
     rRNA            3078..3357
                     /gene="LSU-alpha"
                     /product="large subunit ribosomal RNA"
     gene            3078..3357
                     /gene="LSU-alpha"
     tRNA            3373..3439
                     /partial
                     /gene="tRNA-iMet"
                     /product "transfer RNA-Met"
     gene            3373..3439
                     /gene="tRNA-fMet"
BASE COUNT     1337 a    367 c    493 g   1242 t
ORIGIN
```

```
  Query Match              3.8%;  Score 40;  DB 33;  Length 3439;
  Best Local Similarity   52.4%;  Pred. No. 17;
  Matches   88;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy       5 tgtaaattacatatatttatttttgccgcatttgctaaccgcctattcttttttaaaattt 64
           ||| | ||   | |||||||| ||  |  ||||| |        | ||||      |||||
Db      96 TGTTATTTTTTTTTATTTATATTTTCATATTTTTCATTTGTATTCTTCTAAAATTAATTT 155

Qy      65 taacgattaagaacgcaatttcaatttgcgttctacacaaattaacaattcgtccaaaaag 124
           ||  | |||  |  |||| ||| ||  ||      |||| || || |  |  | | ||
Db     156 AAAATAATAAATAAAAAATTAAATCTTTTTTTAATTGAAAATAAAAAAATAATTAATAAC 215

Qy     125 aggtattaaagcgaagattgcgcatttttttaatccatttttaaaaaaa 172
           ||  ||||| | || ||          | | |  | | ||| ||||
Db     216 AGCGTTTAAACAAAAATAATAATAAAAATAATTTTAATATTATAAAA 263
```

```
RESULT  15
AF160864
LOCUS       AF160864     47296 bp    DNA             INV       06-JAN-2000
DEFINITION  Tetrahymena pyriformis mitochondrial DNA, complete genome.
ACCESSION   AF160864
VERSION     AF160864.1  GI:5306128
KEYWORDS    .
SOURCE      Tetrahymena pyriformis.
  ORGANISM  Mitochondrion Tetrahymena pyriformis
            Eukaryota; Alveolata; Ciliophora; Oligohymenophorea;
            Hymenostomatida; Tetrahymenina; Tetrahymena.
REFERENCE   1  (bases 1 to 47296)
  AUTHORS   Edqvist,J., Burger,G. and Gray,M.W.
  TITLE     Expression of mitochondrial protein coding genes in Tetrahymena
            pyriformis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 47296)
  AUTHORS   Burger,G., Zhu,Y., Littlejohn,T.G., Greenwood,S.J., Schnare,M.N.,
            Lang,B.F. and Gray,M.W.
  TITLE     Complete sequence, gene content and organization of the
            mitochondrial genome of Tetrahymena pyriformis. Comparison with
            Paramecium aurelia mitochondrial DNA
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 47296)
  AUTHORS   Burger,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUN-1999) Departement de Biochimie, Universite de
            Montreal, 2900 Boulevard Edouard-Montpetit, Montreal, Quebec H3T
```

```
            1J4, Canada
FEATURES             Location/Qualifiers
     source          1..47296
                     /organism="Tetrahymena pyriformis"
                     /organelle="mitochondrion"
                     /db_xref="taxon:5908"
     repeat_region   complement(1..62)
                     /note="Signal rep1;  Copy number 2;  Telomeric repeat
                     array;  Array of variable length"
                     /rpt_unit complement(32..62)
     tRNA            complement(111..190)
                     /gene="trnY(gua)"
                     /product="tRNA-Tyr"
                     /anticodon=(pos:155..157,aa:Tyr)
     gene            complement(111..190)
                     /gene="trnY(gua)"
     rRNA            complement(205..484)
                     /gene="rnl_a_1"
                     /note="previously designated rnl_alpha_1"
                     /product="large subunit ribosomal RNA"
     gene            complement(205..484)
                     /gene="rnl_a_1"
     misc_feature    complement(205..2894)
                     /note="Copy number 1;  Repeat encompasses rnl gene region"
                     /note="region: Signal repeq"
     variation       303
                     /note="Variation d16;  Copy number 1;  T in rnl_a_2; same
                     variation seen in different isolate of same strain -
                     compare entries TETMTTRGA (qi|343465) and TETMTRGTG
                     (qi|343463)"
                     /replace="T"
     tRNA            complement(496..575)
                     /gene="trnL(uaa)_1"
                     /note="previously designated trnL(uaa)_L;  Copy number 1"
                     /product="tRNA-Leu"
                     /anticodon=(pos:539..541,aa:Leu)
     gene            complement(496..575)
                     /gene="trnl(uaa)_1"
     rRNA            complement(579..2892)
                     /gene="rnl_b_1"
                     /note="previously designated rnl_beta_1"
                     /product="large subunit ribosomal RNA"
     gene            complement(579..2892)
                     /gene="rnl_b_1"
     variation       1099
                     /note="Variation d15;  Copy number 1;  G in rnl_b_1 of
                     TETMTTRGA (qi|343465); variation seen only in different
                     isolate of same strain"
                     /replace="G"
     variation       1262
                     /note="Variation d14;  Copy number 1;  A in rnl_b_2 of
                     TETMTRGTG (qi|343463); variation seen only in different
                     isolate of same strain"
                     /replace="A"
     variation       1266
                     /note="Variation d13;  Copy number 1;  T in rnl_b_2 of
                     TETMTRGTG (qi|343463); variation seen only in different
                     isolate of same strain"
                     /replace="T"
     variation       2370
                     /note="Variation d12;  Copy number 1;  T in rnl_b_2, same
                     variation seen in different isolate of same strain -
                     compare entries TETMTTRGA (qi|343465) and TETMTRGTG
                     (qi|343463)"
                     /replace="T"
     variation       2619
                     /note="Variation d11;  Copy number 1;  A in rnl_b_2;
                     variation not seen in different isolate of same strain -
                     compare entries TETMTTRGA (qi|343465) and TETMTRGTG
                     (qi|343463)"
                     /replace="A"
     gene            complement(3078..3380)
                     /gene="orf100"
     CDS             complement(3078..3380)
                     /gene="orf100"
                     /note="open reading frame ymf57 (C
                     /codon_start=1
                     /transl_table=4
                     /product="orf100"
                     /protein_id="AAD41914.1"
                     /db_xref="GI:5306129"
                     /translation="MEKNRLIFFFFFFVLSSFYY                    MKNYIIYS
                     GEKFMIFYLKKLIKFLFNSNKFNFLNLIESKYS                      SVSIIII
     gene            complement(3380..4746)
                     /gene="orf444"
     CDS             complement(3380..4746)
                     /gene="orf444"
                     /note="open reading frame ymf58 (C                     ..rate
                     codon; potential ATG initiation c                      ..ire
                     /codon_start=1
                     /transl_table=4
                     /product="orf444"
                     /protein_id="AAD41914.1"
                     /db_xref="GI:5306130"
                     /translation="MKGHIQQWGGIEINQIDVI                      RFSIGLI
                     SGIIWIYTSTESMLAYDFKVWYLWEGLSNFDDS                      SIIAFQLI
                     LSKYLNVLVLNNFLNADWYRMLLHSKFNALIWI                      SFFIYEY
                     FIYFDKSNPDMCILAHLLYIHLIVLFLIFIGFVI                     LFNIIDSI
                     GIVEAFKEIISIDDYIGLVFVVAYVFGVFFFIHA                     MSYYSIFMV
                     LCMFTILTLLVFILAYLKCACKNFNSAVEVVI                       ALWVFI
                     TFISLSHYVAEFFIINSALIGSENQSEGMNFLHF                     IFFIYW
                     HTLILVSSQTIAFTAIVFWLFLFLYIFFIIEKF                      FFIFFIYM
     gene            complement(4499..6261)
                     /gene="orf421"
     CDS             complement(4499..6261)
                     /gene="orf421"
                     /note="open reading frame ymf76 (
                     /codon_start=1
                     /transl_table=4
                     /product="orf421"
                     /protein_id="AAD41915.1"
                     /db_xref="GI:5306131"
                     /translation="MNIIWVFKIKTKKIIFKLL                      INNYKII
                     FIKNIIKIPFYIFIFYFPYKILPKKKFPKDEN                       FFSKIFKYF
                     KIFFFFMMFFNVLSCLFFKSKLINFIVFFF                         FINLSKF
                     KKFFFKKIKSFRIIKYFMKFYKKLSYFQFKHIN                      QFFFFKIHF
                     FIISYLNNNFNIYKIYKYKIKHNYKFIKRIFC                       FFSIFSIRNF
                     YAYTFRNADIFIKSIFIFLNGLQEINPLKYIYF                      IFKKKII
                     PMSFIFYIYWIILFNFVNPFKKFLMISFLFF                        IYFILF
                     MKTKKIIFYLNFFIIFIYLLFLFNWKLIS"
     gene            complement(6278..7106)
                     /gene="rps14"
     CDS             complement(6278..7106)
                     /gene="rps14"
                     /note="ATA initiation codon; AT                        start
                     /codon_start=1
                     /transl_table=4
                     /product="ribosomal protein S14"
                     /protein_id="AAD41916.1"
                     /db_xref="GI:5306132"
                     /translation="MKEFFQLNGYIFNKDTNI                       NYIFKIIYI
                     HNNISYFSAEWKFKSILIFFFKSFNLKNIS                         NIYFGLRHF
                     QQKIWINAWSIYKSNILNLKFFKLEIAKFIYCN                      NLINYIW
                     WLQAKSKFKFMQNFNHNMFKIIGLNSMASGTI                       SSSQFKKN
                     HGFILFYLKFNNSLSAKEKNKIKIVLTEEFI                        FFKIFKFSS
                     "
     misc_feature    complement(7138)
                     /note="RNA 5'-end;  RNA process                        from
                     /note="region: Signal rproc1"
     gene            complement(7210..7565)
                     /gene="rps4"
     CDS             complement(7210..7565)
                     /gene="rps4"
                     /codon_start=1
                     /transl_table=4
                     /product="ribosomal protein S4"
```

```
                     /protein_id="AAD41917.1"
                     /db_xref="GI:5306133"
                     /translation="MGLKSLPLLNKSITGMFWNNVWDSTKLYKAYSIGFLYLNDVTFY
                     FLNENLYYYCIMKTRLTGNEYRGIKGFKQINMNKMKKSWNMKNFYLGKTLPLKSQGWV
                     IVLLNYYSSKKNFLYKAYKSFVFKKLFKSFSKNIRKCNSKTDNYKFKF"
     gene            complement(7574..7870)
                     /gene="rps19"
     CDS             complement(7574..7870)
                     /gene="rps19"
                     /codon_start=1
                     /transl_table=4
                     /product="ribosomal protein S19"
                     /protein_id="AAD41918.1"
                     /db_xref="GI:5306134"
                     /translation="MGKSNWFLNTIKSIMSKIFNKSIAIKKKKISLIMNKSSTIPL
                     SLFSNLFFINKCFKYKFLKSHYRIGKRIGFLLTKKFNVRIKKSKKSKSKSFRR"
     gene            complement(7879..8667)
                     /gene="rpl2"
     CDS             complement(7879..8667)
                     /gene="rpl2"
                     /codon_start=1
                     /transl_table=4
                     /product="ribosomal protein L2"
                     /protein_id="AAD41919.1"
                     /db_xref="GI:5306135"
                     /translation="MFLFNYFFFTPTLFFFFSIYFMSFFIFFNNLIFIFNNKSGESN
                     KGNIILRHRGGNIYNKQINLNINRNQINKIGTTSSLNYINKNGCFIGLIKYSNNSLSY
                     IKISNGLFIGNYIKSIDKFINISFNFKNIIGSFIILNLAFKHSVFSNIISIDKNKSQY

  Query Match             3.8%;  Score 40;  DB 34;  Length 47296;
  Best Local Similarity   52.4%;  Pred. No. 15;
  Matches   88;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy       5 tgtaaattacatatatttatttttqccgcatttttqctaacccqcctattcttttttaaaattt 64
           ||| | ||   | |||||||| ||  |  ||||| |        | ||||       |||||
Db   43856 TGTTATTTTTTTTTATTTATATTTTCATATTTTTCATTTGTATTCTTCTAAAATTAATTT 43915

Qy      65 taacqattaaqaacqcaattcaatttqccgttctacacaaattaacaattcqttccaaaaq 124
             ||  | |||  |  |||| ||| ||   ||      |||| || || |  | | ||
Db   43916 AAAATAATAAATAAAAAATTAAATCTTTTTTTAATTGAAAATAAAAAAATAATTAATAAC 43975

Qy     125 aqqtatttaaqccaaqatttqqcatttttttaatccattttttaaaaaa 172
           ||   ||||| | || ||           | | |  | | ||| ||||
Db   43976 AGCGTTTTAAACAAAAATAATAATAAAAATAATTTTAATATTATAAAA 44023


Search completed: August 25, 2000, 22:56:56
Job time: 16388 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2000, 20:19:24 ; Search time 54.3 Seconds
(without alignments)
4824.346 Million cell updates/sec

Title: US-09-257-585-1
Perfect score: 1047
Sequence: 1 gatctgtaaattacatatat.....atggaggcgaccgccggatcc 1047

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	39	3.7	5259	1	X12952	Enterococcus faeca
2	38.6	3.7	110000	1	X20248_06	Continuation (7 of
3	38.6	3.7	111309	1	X20250	Borrelia burgdorfe
c 4	37.8	3.6	3891	1	V57327	Hormone-regulated
c 5	37.8	3.6	3891	1	V54208	Repro-PC-1.0 cDNA
c 6	37.6	3.6	2058	1	V07560	Neocallimastix pat
c 7	37	3.5	600	1	V86185	EST clone J635, Ne
c 8	36	3.4	3800	1	X28141	CCoAOMT promoter,
c 9	35.8	3.4	3055	1	V74608	Staphylococcus aur
10	35.6	3.4	647	1	X47507	Human secreted pro
11	35.4	3.4	1562	1	T84168	DNA encoding 2 elo
12	35.4	3.4	15598	1	V74393	Staphylococcus aur
c 13	35.2	3.4	110000	1	X20248_00	Borrelia burgdorfe
14	33.2	3.2	4200	1	T78912	Spinocerebellar at
c 15	32.8	3.1	110000	1	V21209_04	Continuation (5 of
c 16	32.8	3.1	110000	1	V21209_05	Continuation (6 of
17	32.6	3.1	110000	1	V21209_03	Continuation (4 of
18	32.4	3.1	6285	1	X13352	Enterococcus faeca
c 19	32.2	3.1	5563	1	X08941	GATA-1 promoter re
20	32.2	3.1	9062	1	V74396	Staphylococcus aur
c 21	32.2	3.1	10391	1	X13070	Enterococcus faeca
c 22	32.2	3.1	31096	1	V74370	Staphylococcus aur
c 23	32	3.1	731	1	Q31693	RsaI restriction f
24	32	3.1	1887	1	X13315	Enterococcus faeca
c 25	32	3.1	9542	1	X20260	Borrelia burgdorfe
c 26	31.8	3.0	4526	1	V72245	G. max SBP1 DNA 5'
27	31.8	3.0	18359	1	X20255	Borrelia burgdorfe
28	31.8	3.0	110000	1	V21209_06	Continuation (7 of
29	31.8	3.0	110000	1	V21209_07	Continuation (8 of
30	31.8	3.0	110000	1	X20248_00	Borrelia burgdorfe
c 31	31.6	3.0	6360	1	Q11378	Genes encoding T.
32	31.4	3.0	858	1	Q55451	Fgs28 protein codi
33	31.4	3.0	858	1	Q87432	P. gallinaceum Pgs
34	31.4	3.0	3661	1	[illegible]	[illegible]
35	31.4	3.0	5737	1	Q58054	[illegible]
36	31.4	3.0	5737	1	Q74053	[illegible]
37	31.4	3.0	110000	1	X20248_04	[illegible]
c 38	31.4	3.0	110000	1	X20248_07	[illegible]
c 39	31.4	3.0	116277	1	X20249	[illegible]
c 40	31.2	3.0	442	1	V74852	[illegible]
41	31.2	3.0	1095	1	I91545	[illegible]
c 42	31.2	3.0	1422	1	V60073	[illegible]
43	31.2	3.0	2492	1	Q79480	[illegible]
44	31.2	3.0	3480	1	V42046	[illegible]
c 45	31.2	3.0	4599	1	N00472	[illegible]

ALIGNMENTS

RESULT 1
X12952
ID X12952 standard; DNA; 5259 BP.
AC X12952;
DT 19-MAR-1999 (first entry)
DE Enterococcus faecalis genome contig SEQ ID NO:15.
KW Enterococcus faecalis; contig; detection; Enteroc
KW vaccine; attenuation; computer readable medium; d
OS Enterococcus faecalis.
PN WO9850555-A2.
PD 12-NOV-1998.
PF 04-MAY-1998; U08985.
PR 14-NOV-1997; US-066005.
PR 06-MAY-1997; US-044031.
PR 16-MAY-1997; US-046655.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Dillon PJ, Kunsch CA;
DR WPI; 99-045171/04.
PT New isolated Enterococcus faecalis polynucleotide
PT - used to develop products for the detection of E
PT use in vaccines for prevention or attenuation of
PT infection.
PS Claim 1; Page 313-316; 2084pp; English.
CC A computer readable medium has been developed wh
CC 982 nucleotide sequences isolated from the Entero
CC X12938 to X14919 represent these nucleotide seque
CC nucleotide sequences, also known as contigs. The
CC can identify fragments of the Enterococcus faeca
CC commercial importance. The products can be used f
CC of Enterococcus faecalis in samples. They can al
CC diagnosing Enterococcal infection in an animal a
CC progression of disease, and for identifying agent
CC modulate the growth or pathogenicity of Enteroco
CC another related organism, in vivo or in vitro. T
CC polypeptides encoded by the Enterococcus faecali
CC can be used in vaccines to prevent or attenuate
CC infection.
SQ Sequence 5259 BP; 1778 A; 885 C; 1094

Query Match 3.7%; Score 39; DB 1;
Best Local Similarity 49.3%; Pred. No. 0.11;
Matches 102; Conservative 0; Mismatches 105

Qy 23 atttttgccgcatttttgctaaacgcgtattgttttttaaaaattt 72
Db 2275 AGTTTCCAGCCACATTTATCAACAAACCATCCAATCTAAAAAGT 2324

Qy 83 ttcaattttgcgttttctaccgcaaattaacaattcgtcccaaaaa 132
Db 2325 TTTTAAAGGTccCTACAACAAAAGAAACAGTTCGTTGTAAT 2374

Qy 143 ttggcatttttttatccgtttttttaaaaaaatacatctgctt 192
Db 2375 TTGGCACATTTA AcAAc cAATTAcAAAAGAAATTccTGTTc 2424

```
Qy     209 gataccccgctagcatagccttcgttta 229
           || |    || | |   ||||| | |
Db    2455 GGCAAAAAGCCAACTGCGCCTTATTAA 2481


RESULT   2
X20248_06
Continuation (7 of 10) of X20248 from base 600001 (Borrelia burgdorferi polynucleotide s
WP Sequence split into 10 fragments  LOCUS X20248 Accession X20248
WP      Fragment Name       Begin      End
WP      X20248_00                1   110000
WP      X20248_01           100001   210000
WP      X20248_02           200001   310000
WP      X20248_03           300001   410000
WP      X20248_04           400001   510000
WP      X20248_05           500001   610000
WP      X20248_06           600001   710000
WP      X20248_07           700001   810000
WP      X20248_08           800001   910000
WP      X20248_09           900001   910715


  Query Match              3.7%;  Score 38.6;  DB 1;  Length 110000;
  Best Local Similarity   58.1%;  Pred. No. 0.51;
  Matches   68;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy      17 atatttatttttgccgcatttttgctaaccgccatattcttttttaaaatttttaacgattaaga 76
           | ||||||||||   | |||||   ||   | ||  ||   | |   ||      |||| |
Db   72216 AAATTTATTTTTATGAATTTTTTGAAAAAATTTTTATTGCCAGACAATTTGAATTTAACA 72275

Qy      77 acgcaattcaatttttgcgttctacacaaaattaacaatttgtccaaaaagaggtatttta 133
           |   || | | || ||| || || ||||| ||    ||   ||||| |||||| ||
Db   72276 ATAAAAATTATTTATGCTTTTTATACAAAGAAAATCATCAGACAAAAAAGGTATATA 72332


RESULT   3
X20250
ID   X20250 standard; DNA; 111309 BP.
AC   X20250;
DT   04-MAY-1999  (first entry)
DE   Borrelia burgdorferi polynucleotide sequence #3.
KW   Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
KW   epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
KW   infection; diagnosis; characterisation; detection; ds.
OS   Borrelia burgdorferi.
PN   WO9858948-A1.
PD   30-DEC-1998.
PF   18-JUN-1998; 98WO-US012764.
PR   03-SEP-1997; 97US-057483.
PR   20-JUN-1997; 97US-050359.
PR   22-JUL-1997; 97US-053344.
PR   22-JUL-1997; 97US-053377.
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (MEDI-) MEDIMMUNE INC.
PI   Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO,
PI   White OR;
DR   WPI; 99-081217/07.
PT   New isolated Borrelia burgdorferi nucleic acids - used to develop
PT   products for the detection, diagnosis, characterisation, prevention
PT   and therapy of infections, particularly Lyme disease
PS   Claim 1; Page 738-800; 1128pp; English.
CC   X20248 to X20402 represent polynucleotide sequences isolated from
CC   Borrelia burgdorferi (Bb). Products derived from Bb can be used for
CC   the detection, diagnosis, characterisation, prevention and therapy of
CC   Bb infections, e.g. Lyme disease. They can also be used for the
CC   production of biosynthetic products, e.g. enzymes. Borrelia belongs
CC   to a family of motile, spiral-shaped bacteria called Spirochetes.
CC   Spirochetes are pathogenic in humans and Borrelia causes epidemic and
CC   endemic relapsing fever, and Lyme borreliosis, more commonly known as
CC   Lyme disease.
SQ   Sequence   111309 BP;    35956 A;    13151 C;    19075 G;    43117 T;


  Query Match              3.7%;  Score 38.6;  DB 1;  Length 111309;
  Best Local Similarity   58.1%;  Pred. No. 0.52;
  Matches   68;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy      17 atatttatttttgccgcatttttgctaaccgccatattcttttttaaaatttttaacgattaaga 76
           | ||||||||||   | |||||   ||   | ||  ||   | |   ||      |||| |
Db   82840 AAATTTATTTTTATGAATTTTTTGAAAAAATTTTTATTGCCAGACAATTTGAATTTAACA 82899

Qy      77 acgcaattcaatttttgcgttctacacaaaattaacaatttgtccaaaaagaggtatttta 133
           |   || | | || ||| || || ||||| ||    ||   ||||| |||||| ||
Db   82890 ATAAAAATTATTTATGCTTTTTATACAAAGAAAATCATCAGACAAAAAAGGTATATA 82946


RESULT   4
V57327/A
ID   V57327 standard; cDNA; 3891 BP.
AC   V57327;
DT   21-DEC-1998  (first entry)
DE   Hormone-regulated Repro-PC-1.0 gene.
KW   Repro-PC-1.0; prostate cancer; LNCaP; hormone-regulated gene;
KW   human; synaptotagmin; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             151. .1428
FT                   /*tag= a
PN   WO9839661-A1.
PD   11-SEP-1998.
PF   06-MAR-1998; 98WO-US004519.
PR   15-MAY-1997; 97US-047811.
PR   07-MAR-1997; 97US-041246.
PA   (REPR-) REPROGEN INC.
PI   French CK, Yamamoto KK;
DR   WPI; 98-506479/43.
DR   P-PSDB; W75782.
PT   Identification of hormone-regulated traits, e.g. in cancers - by
PT   exposing grafts of biological material to different hormonal
PT   environments in animals of different reproductive states
PS   Example 1; Page 59-62; 85pp; English.
CC   This nucleotide sequence represents a hormone-regulated gene from
CC   human prostate cancer cells.  In order to isolate sequences that
CC   are over expressed in male LNCaP tumours, a male-LNCaP-specific
CC   probe was generated by 3 rounds of subtractive hybridisation with
CC   female LNCaP tumour cDNA.  The probe was used to perform a primary
CC   screen of a lambda-ZAP male LNCaP tumour cDNA library.  Positive
CC   plaques were subjected to secondary and then tertiary screens using
CC   male- and female-specific probes to isolate partial clone
CC   Repro-PC-1.0.  Subsequent screening of the male-LNCaP tumour
CC   library and RACE-PCR yielded a sequence containing a single 1275
CC   open reading frame encoding 425 amino acids (see W75782).  The
CC   encoded protein has regions of homology to the C2 regulatory
CC   domain of calcium-dependent isoforms of protein kinase C and to
CC   isoforms of synaptotagmin.  The gene was localised to chromosome
CC   18.  The invention provides methods of identifying hormone-regulated
CC   traits in a cell.  The methods involve cultivating the cell as a
CC   graft in 2 different hormonal environments and determining whether
CC   expression of the trait differs in the 2 grafts.  The methods can
CC   be used to identify hormonally-regulated traits and hormonally
CC   regulated genes for use as targets for therapeutic intervention in
CC   disease states, particularly cancers.
SQ   Sequence   3891 BP;    1229 A;    671 C;    788 G;    1203 T;


  Query Match              3.6%;  Score 37.8;  DB 1;  Length 3891;
  Best Local Similarity   53.8%;  Pred. No. 0.22;
  Matches   78;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy      48 tattcttttttaaaatttttaacgattaaqaacgcaattcaatttttgcgttctacacaaaatt 107
           || | ||  | || |||| ||  | || ||| |  |  | |||        | |   |||
Db    3474 TAGTTTTAGTCAACTTTTCACACTCAAAAACACTTTAAATTTTAAAAGCTAAAAACTATT 3415

Qy     108 aacaattcgtccaaaaagaqgtatttaaqcgaaqatttqqcatttttttttaatccatttttta 167
           |||| |  ||||    |  | |  ||  |    ||     |||  ||  ||||| || |
Db    3414 TACAAATTATCCATTTTAATTTTTAAAAATGTTTGTTGTTGGTTTCATTGTTCCATGTTAA 3355
```

```
Qy      168 aaaaaatacatctgctttaacccac 192
            ||||||| | ||| || |   || ||
Db     3354 AAAAAAAAAATCGGCCTCGTCCAAC 3330


RESULT   5
V54208/c
ID   V54208 standard; cDNA; 3891 BP.
AC   V54208;
DT   11-JAN-1999  (first entry)
DE   Repro-PC-1.0 cDNA encoding prostate cancer-specific marker.
KW   Repro-PC 1.0; prostate cancer; marker; synaptotagmin; human;
KW   diagnosis; therapy; vaccine; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             151. .1428
FT                   /*tag= a
PN   WO9839447-A1.
PD   11-SEP-1998.
PF   06-MAR-1998; U04488.
PR   15-MAY-1997; US-047811.
PR   07-MAR-1997; US-041246.
PA   (REPR-) REPROGEN INC.
PI   French CK, Schneider PA, Yamamoto KK;
DR   WPI; 98-506363/43.
DP   P-PSDB: W74584.
PT   New isolated Repro-PC-1.0 polynucleotide(s)   are used to develop
PT   products for the detection and prophylactic and therapeutic
PT   treatment of prostate cancer
PS   Claim 4; Page 64-67; 87pp; English.
CC   This cDNA sequence codes a prostate cancer-specific marker, termed
CC   Repro-PC-1.0 (see W74584), which represents a novel human brain
CC   synaptotagmin isoform that may function in exocytosis and
CC   endocytosis pathways. Subtractive hybridisation was used to
CC   isolate Repro-PC-1.0 cDNA from a male LNCaP tumour cDNA library.
CC   An overlapping clone (PS5-1) was subsequently obtained from the
CC   library using a probe containing 5' sequences of Rep-PC-1.0, and
CC   the complete coding region was determined by sequencing PS5-1 and
CC   an overlapping RACE-PCR derived 5' end cDNA clone. The
CC   Repro-PC-1.0 gene localises to chromosome 18. The invention
CC   provides Repro-PC-1.0 polynucleotides, including probes and
CC   primers, antisense sequences useful in the treatment of prostate
CC   cancer, as well as a polynucleotide vaccine for eliciting an
CC   immune response against Repro-PC-1.0. Also claimed are methods
CC   for detecting Repro-PC-1.0 polynucleotides using the probes and
CC   primers, a method of inhibiting Repro-PC.10 expression in a cell
CC   using the antisense sequence, methods for diagnosing prostate
CC   cancer and for detecting prostate cancer cells in a subject, for
CC   following the progress of prostate cancer, for detecting a
CC   chromosomal translocation of a Repro-PC-1.0 gene, and for
CC   detecting polymorphic forms of Repro-PC-1.0.
SQ   Sequence   3891 BP;    1229 A;    671 C;    788 G,    1203 T;


  Query Match              3.6%;  Score 37.8;  DB 1;  Length 3891;
  Best Local Similarity   53.8%;  Pred. No. 0.22;
  Matches   78;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy       48 tattctttttaaaatttaacgattaagaacgcaatttcaatttttgcgttctacacaaaatt 107
            || | ||  | || |||| ||  | || ||| |  |  | |||       | |   |||
Db     3474 TAGTTTTACTCAACTTTTCACACTCAAAAACACTTTAAATTTTAAAAGCTAAAAACTATT 3415

Qy      108 aacaattcgtccaaaaqaqqtatttaaqcqaaqatttqqcatttttttaatccatttttta 167
             |||| |  ||||    |  | |  || |    ||     ||| ||  ||||| || |
Db     3414 TACAAATTATCCATTTTAATTTTAAAAATGTTTGTTGTTGGTTTCATTGTTCCATGTTAA 3355

Qy      168 aaaaaatacatctgctttaacccac 192
            ||||||| | ||| || |   || ||
Db     3354 AAAAAAAAAATCGGCCTCGTCCAAC 3330


RESULT   6
```

```
V07560/c
ID   V07560 standard; DNA; 2058 BP.
AC   V07560;
DT   09-NOV-1998  (first entry)
DE   Neocallimastix patriciarum endo-xylanase gene.
KW   Endo-xylanase; xynC gene; transgenic plant; Brassic
KW   oilseed rape; canola; feedstuff; ds.
OS   Neocallimastix patriciarum strain 27.
FH   Key             Location/Qualifiers
FT   CDS             301. .1758
FT                   /*tag= a
FT   sig_peptide     301. .360
FT                   /*tag= b
FT                   /note= "putative signal peptide"
FT   mat_peptide     361. .1755
FT                   /*tag= c
PN   CA2190194-A.
PD   13-MAY-1998.
PF   13-NOV-1996; 190194.
PR   13-NOV-1996; CA-190194.
PA   (CHEN/) CHENG K.
PA   (FORS/) FORSBERG C W.
PA   (HUYY/) HU Y.
PA   (LIUJ/) LIU J H.
PA   (MOLO/) MOLONEY M M.
PA   (SELI/) SELINGER L B.
PI   Cheng K, Forsberg CW, Hu Y, Liu JH, Moloney MM,
PI   Selinger LB;
DR   WPI; 98 428424/37.
DR   P-PSDB; W65462.
PT   New DNA encoding Neocallimastix patriciarum xylana
PT   produce transgenic plant and as feed supplement
PS   Claim 2; Page 32-34; 49pp; English.
CC   This isolated genomic DNA clone, designated pNspX
CC   an endo-xylanase (see W65462) of Neocallimastix pa
CC   27   Using Congo Red detection, a N. patriciarum
CC   library was screened for lambda clones possessing
CC   activity. A positive clone carrying a 6.5 kb Eco
CC   identified. Plasmid DNA from this clone was intr
CC   coli, and the gene (xynC) encoding the observed x
CC   in recombinant clones was identified by deletion
CC   sequence analysis. The invention provides host c
CC   with the above DNA. Recombinant enzyme isolated
CC   cells showed optimal activity at pH 6.2 and 40 de
CC   value of 0.89 mg/l for oat spelt xylan hydrolysis
CC   can be used to design probes for use in isolating
CC   from other anaerobic fungi. It has been used to
CC   oleosin-xynC expression construct encoding an ole
CC   fusion protein which retains xylanase activity.
CC   plants transformed with this expression construct
CC   fusion protein in the oil-body membrane of seeds
CC   prepared from such plants retains xylanase activ
CC   ideal animal feed supplement. E. coli DH5alpha
CC   the xynC gene is deposited at ATCC 98249.
SQ   Sequence   2058 BP;    692 A;    348 C;    488 G


  Query Match              1.6%;  Score 37.6;  DB 1;
  Best Local Similarity   90.5%;  Pred. No. 0.19;
  Matches   70;  Conservative    0;  Mismatches   54;

Qy        2 atctgtaaattacatatatttatttgccgcatttgctaacc
            || | |   ||  |||||||| ||||     |||||  |||
Db     2031 ATATTTTTTTTTATAIATAITTTTTTTTATAATTTTATTAAAA

Qy       62 ttttaacgattaagaacgcaatttcaatttgcgttctacacaa
            | ||||  |||||  |     || || ||| | ||
Db     1971 TATTAAAIAITAAA!AAAIIGIIIAAAIITACATTIACAIIIA

Qy      122 aaqa 125
            || |
Db     1911 AAAA 1908
```

```
RESULT   7
V86185/c
ID   V86185 standard; cDNA; 600 BP.
AC   V86185;
DT   27-APR-1999  (first entry)
DE   EST clone J635.
KW   Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW   tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW   chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW   receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
OS   Homo sapiens.
PN   WO9845435-A2.
PD   15-OCT-1998.
PF   10-APR-1998; 98WO-US006954.
PR   10-APR-1997; 97US-0835913.
PA   (GEMY ) GENETICS INST INC.
PI   Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI   Racie LA, Spaulding V, Treacy M;
DR   WPI; 99-070076/06.
PT   New polynucleotides encoding human secreted proteins - derived from
PT   e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT   ovary, pituitary, retina and colon cDNA libraries
PS   Claim 1; Page 150; 643pp; English.
CC   This sequence represents an expressed sequence tag (EST), and is a
CC   polynucleotide of the invention. The polynucleotides of the invention are
CC   all secreted EST sequences isolated from a variety of human tissue
CC   sources. The EST sequences and proteins encoded by them are predicted to
CC   have useful biological activities which would make them suitable for
CC   treating, preventing or ameliorating medical conditions in humans and
CC   animals, although no supporting data is given. Suggested activities
CC   include nutritional activity, immune stimulating or suppressing activity,
CC   haematopoiesis regulating activity, tissue growth activity,
CC   activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC   activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC   activity. The EST sequences are also stated to be useful for gene
CC   therapy.
SQ   Sequence   600 BP;    151 A;    106 C;    94 G;    127 T;


  Query Match              3.5%;  Score 37;  DB 1;  Length 600;
  Best Local Similarity   17.0%;  Pred. No. 0.17;
  Matches   28;  Conservative   7;  Mismatches   61;  Indels    0;  Gaps    0;

Qy      34 ttttgctaaccgcctattctttttaaaatttaaacgattaagaaccgcaattcaatttttgc 93
           ::|: :   :  ::  :: |:::|:|:::: :||   :: : :| ::  ..|.|  ..|
Db     394 YYTYYYCGYSKYYYCYYYKTKKTTWAMMWWAWAAWYCYWYMMMAARSCCYWCMAAMYKGG 335

Qy      94 gttctaacacaaattaacaattcgtccaaaaaqaggtatttaaqccaaqatttqqcatttttt 153
           || :|  :   : :: : :::    |: : ::: :::   :   : ::|:::::::::|
Db     334 YTTTWAWYYYCYYMMMAASWYMAKCWAKCWWMKCGWYYSYYYYCYAAKKIKSMWYYYYT 275

Qy     154 ttaatccatttttaaaaaaatacatctgctttaacccaccttttgc 198
           ::::: :||: : :|:::: | :::| : |::|    :| || |
Db     274 GKWWYSCTTTYMWYIAWMMMAAAWYTTYMATKAAKWTYCCCTIC 230


RESULT   8
X28141/c
ID   X28141 standard; DNA; 3800 BP.
AC   X28141;
DT   11-JUN-1999  (first entry)
DE   CCoAOMI promoter, PtCCoAOMT1.
KW   CCoAOMI promoter; Poplar; caffeoyl CoA O-methyltransferase promoter;
KW   PtCCoAOMT1; plant specific promoter; xylem ray cell;
KW   lignin level alteration; ss.
OS   Populus trichocarpa.
PN   WO9909188-A2.
PD   25-FEB-1999.
PF   27-JUL-1998; 98WO-EP04988.
PR   13-AUG-1997; 97EP-202507.
PA   (VLAA-) VIAAMS INTERUNIVERSITAIR INST BIOTECHNOG.
PI   Boerjan W, Chen C, Meyermans H, Van Montagu M;
DR   WPI; 99-190169/16.
PT   Poplar caffeoyl-CoA-O-methyltransferase (CCoAOMI) promoter - to
PT   provide transgenic plants having altered lignin level and/or
PT   composition
PS   Claim 1; Fig 1b; 69pp; English.
CC   This sequence represents the Poplar (Populus trichocarpa)
CC   caffeoyl-CoA-O-methyltransferase (CCoAOMT) promoter, PtCCoAOMT1, of the
CC   invention. The promoter is specific for plant or tree living vessel
CC   elements and/or xylem adjacent ray cells. The promoter is used in genetic
CC   modification to provide transgenic plants having altered lignin level
CC   and/or composition, which improves digestibility of forages and pulping
CC   properties of trees.
SQ   Sequence   3800 BP;    1225 A;    668 C;    701 G;    1206 T;


  Query Match              3.4%;  Score 36;  DB 1;  Length 3800;
  Best Local Similarity   49.5%;  Pred. No. 0.73;
  Matches   93;  Conservative    0;  Mismatches   95;  Indels    0;  Gaps    0;

Qy      42 acggcctattcttttaaaatttaacgattaagaaccgcaattcaattttgcgttctaca 101
           | | |  ||  |||||||||  | ||  | ||  |    ||  || ||    ||   | |
Db     800 ATCACAAATGATTTTAAAAAATAAAAAAATATTATTTTAAAGCATTTCCAAGTAAAAAA 741

Qy     102 caaattaacaattcgtccaaaaaqaggtatttaagccaaqatttqqcatttttttaatcca 161
           ||  || | ||      |  ||  |  | || ||| ||| |||    |  | |||    ||
Db     740 CACTTTGAAAAGCAACCATAACCATTTTTTCAAGTGAACATTAATAAAATATTTTTCACA 681

Qy     162 tttttaaaaaaaatacatctgctttaacccaccttttgccaaggqtaccccggctaqcatagc 221
           | |||   ||||||  |   |    |  | |    ||  ||| |||      | | ||
Db     680 TGTTTTCCAAAATATAATAGAAACATTCGAAAGCTGGGAAGTGTATTTCAATTGAATTTT 621

Qy     222 cttggtta 229
           ||  |||
Db     620 TTTCTTTA 613


RESULT   9
V74608/c
ID   V74608 standard; DNA; 3055 BP.
AC   V74608;
DT   16-MAR-1999  (first entry)
DE   Staphylococcus aureus contig SEQ ID #297.
KW   Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW   cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW   skin infection, surgical wound infection, scalded skin syndrome;
KW   toxic shock syndrome; ds.
OS   Staphylococcus aureus.
FH   Key             Location/Qualifiers
FT   misc_feature    841. .900
FT                   /*tag= a
FT                   /note= "these bases represent a line of missing text in
FT                   the sequence listing in the specification. They
FT                   are included to maintain the nucleotide numbering
FT                   given in the specification for this DNA sequence"
FT   misc_feature    2641. .2700
FT                   /*tag= b
FT                   /note= "these bases represent a line of missing text in
FT                   the sequence listing in the specification. They
FT                   are included to maintain the nucleotide numbering
FT                   given in the specification for this DNA sequence"
PN   EP-786519-A2.
PD   30-JUL-1997.
PF   07-JAN-1997; 100117
PR   05-JAN-1996; US-009861.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI   Rosen CA;
DR   WPI; 97-374922/35.
PT   Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT   stored on computer readable medium and used in the production of
PT   anti-S.aureus vaccines
PS   Claim 1; Page 1123-1125; 3271pp; English.
```

CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
SQ Sequence 3055 BP; 1006 A; 500 C; 423 G; 1006 T;

Query Match 3.4%; Score 35.8; DB 1; Length 3055;
Best Local Similarity 52.4%; Pred. No. 0.76;
Matches 76; Conservative 1; Mismatches 68; Indels 0; Gaps 0;

Qy 20 tttatttttgccgcattttgctaaccgcctattcttttttaaaatttaacgattaagaaacg 79
Db 1410 TGTTTTTAGTTGTATTTKGTTGGCAACCAATTCATTTTTATGCCTTAGCTATTAAACGTA 1351

Qy 80 caattcaatttttgcgttctacacaaattaacaattcgtccaaaagagggtatttaagcgaa 139
Db 1350 AAGATGAATATGCACTTGCAAATATTCCAATGTTACCATCAGTTAAGGGCTTTAAACGTA 1291

Qy 140 gatttqqcatttttttaatccatt 164
Db 1290 CACGTGTCAGTATGTTTATCTGGTT 1266

RESULT 10
X37507
ID X37507 standard; cDNA; 647 BP.
AC X37507;
DT 06-JUL-1999 (first entry)
DE Human secreted protein cDNA fragment containing gene 57.
KW Human; secreted protein; treatment; prevention, protein therapy; AIDS;
KW gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
KW developmental abnormality; fetal deficiency; blood disorder; leukemia;
KW immune system disease; autoimmune disease, hepatic disease; lymphoma;
KW renal disease; inflammation; allergy; Alzheimer's disease; schizophrenia;
KW cognitive disorder; prostate disease; skeletal; cardiac; muscle disorder;
KW pulmonary disorder; transplant rejection; osteoclast; osteoporosis;
KW arthritis; malignancy; digestive; endocrine; infection; ss.
OS Homo sapiens.
PN WO9918208-A1.
PD 15-APR-1999.
PF 01-OCT-1998; U20775.
PR 02-OCT-1997; US-060884.
PR 02-OCT-1997; US-060833.
PR 02-OCT-1997; US-060836.
PR 02-OCT-1997; US-060837.
PR 02-OCT-1997; US-060838.
PR 02-OCT-1997; US-060839.
PR 02-OCT-1997; US-060843.
PR 02-OCT-1997; US-060862.
PR 02-OCT-1997; US-060866.
PR 02-OCT-1997; US-060874.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Carter KC, Duan DR, Endress GA, Feng P, Ferrie AM,
PI Florence KA, Greene JM, Janat F, Lafleur DW, Ni J,
PI Rosen CA, Ruben SM, Shi Y, Young P, Yu G,
DP WPI; 99-264022/22.
DR P-PSDB; Y07908.
PT New isolated human genes and the secreted polypeptides they encode
PS Claim 1a; Page 257; 368pp; English.
CC This invention describes novel isolated human genes
CC proteins they encode. The products of the invention
CC preventing, treating or ameliorating medical condit
CC or gene therapy. Also pathological conditions can b
CC determining the amount of the new polypeptides in a
CC determining the presence of mutations in the new po
CC Specific uses are described for each of the 101 pol
CC which tissues they are most highly expressed in, an
CC products for the diagnosis or treatment of cancer,
CC neurodegenerative disorders, developmental abnormal
CC deficiencies, blood disorders, leukemias, diseases
CC autoimmune diseases, hepatic and renal disease, lym
CC allergies, Alzheimer's and cognitive disorders, sch
CC disease, skeletal or cardiac muscle disorders, pul
CC transplant rejection, disorders involving osteoclas
CC osteoporosis, arthritis or malignancies, digestive,
CC infections and AIDS. The human secreted proteins c
CC represented in Y07852-Y07993 and the encoding nucl
CC represented in X37451-X37562.
SQ Sequence 647 BP; 215 A; 122 C; 117 G;

Query Match 3.4%; Score 35.6; DB 1;
Best Local Similarity 52.7%; Pred. No. 0.46;
Matches 77; Conservative 0; Mismatches 69;

Qy 11 ttacatatatttatttttqccqcatttqctaaccgcctattctt
Db 10 TTTCATATATTTTTTTCTCATCTTTTTGCCTGTTACTTATATGTAA

Qy 71 ttaagaacgcaattcaatttqcgttctacacaaattaacaatt
Db 70 TTGAGATCTTGTTATATATTTTCATTTGTTGCTTTATAACCATT

Qy 131 ttaaqccqaaqatttqqccatttttttta 156
Db 130 TTTAATTAAGGTCTGGAATTTTTTTA 155

RESULT 11
T84168
ID T84168 standard; DNA; 1562 BP.
AC T84168;
DT 14-SEP-1998 (first entry)
DE DNA encoding 2 elongation factor G proteins.
KW Staphylococcus aureus protein; ribozyme; antisens
KW Staphylococcal gene; regulatory element; bacteri
KW vaccine; Staphylococcal infection; food poisonin
KW toxic shock syndrome; ss.
OS Staphylococcus aureus.
FH Key Location/Qualifiers
FT CDS 618. .1097
FT /*tag= a
FT CDS 1190. .1542
FT /*tag= b
PN WO9730070-A1.
PD 21-AUG-1997.
PF 19-FEB-1997; U02318.
PR 20-FEB-1996; US-011888.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PI Black MT, Burnham MK, Rodgers JR, Knowles DJC, N
PI Pratt JM, Reichard RW, Rosenberg M, Ward JM;
DR WPI; 97-424969/39.
DR P-PSDB; W28265, W28266.
PT Novel polypeptide(s) from Staphylococcus aureus
PT to isolate antimicrobial compounds, and in vacc
PT aureus infection
PS Claim 9; Page 929-930; 989pp; English.
CC The present sequence encodes 2 Staphylococcus a
CC based on homology with Anacystis nidulans and M
CC proteins, are believed to be elongation factor
CC sequence was obtained from a library of clones

```
CC   Escherichia coli.  The DNA sequence can be used in the construction
CC   of ribozymes and antisense sequences to control the expression of
CC   Staphylococcal genes.  The DNA sequence is also useful as a source of
CC   regulatory elements for the control of bacterial gene expression.
CC   The encoded protein may be used to produce vaccines to enable a host
CC   to produce specific antibodies with antibacterial action.  These vaccines
CC   and antibodies would protect a host against invasion by S. aureus, and
CC   conditions relating to Staphylococcal infection, e.g. Staphylococcal food
CC   poisoning, scalded skin syndrome, and toxic shock syndrome.
SQ   Sequence   1562 BP;    517 A;    293 C;    329 G;    423 T;


  Query Match             3.4%;  Score 35.4;  DB 1;  Length 1562;
  Best Local Similarity   61.3%;  Pred. No. 0.76;
  Matches   57;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy      94 gttctacacaaattaacaattcgtccaaaaagaggtatttaagcgaagatttggcattttt 153
           | || |  | ||||| ||||| || |  |||| | |||  |   ||   ||| | |  ||
Db     951 GCTCCAATCCAATTACCAATTGGTGCGGAAGACGAATTCGAAGCAATCATTGACTTAGTT 1010

Qy     154 ttaatccatttttaaaaaaatacatctgcttta 186
             |||  | | ||  ||| |||||  || ||||
Db    1011 GAAATGAAATGTTTCAAATATACAAATGATTTA 1043


RESULT  12
V74393
ID   V74393 standard; DNA; 15598 BP.
AC   V74393;
DT   16-MAR-1999  (first entry)
DE   Staphylococcus aureus contig SEQ ID #82.
KW   Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW   cellulitis; eyelid infection; food poisoning; osteomyelitis, therapy;
KW   skin infection; surgical wound infection; scalded skin syndrome;
KW   toxic shock syndrome; ds.
OS   Staphylococcus aureus.
FH   Key             Location/Qualifiers
FT   misc_feature    481. .540
FT                   /*tag= a
FT                   /note= "these bases represent a line of missing text in
FT                         the sequence listing in the specification. They
FT                         are included to maintain the nucleotide numbering
FT                         given in the specification for this DNA sequence"
FT   misc_feature    2281. .2340
FT                   /*tag= b
FT                   /note= "these bases represent a line of missing text in
FT                         the sequence listing in the specification. They
FT                         are included to maintain the nucleotide numbering
FT                         given in the specification for this DNA sequence"
FT   misc_feature    4081. .4140
FT                   /*tag= c
FT                   /note= "these bases represent a line of missing text in
FT                         the sequence listing in the specification. They
FT                         are included to maintain the nucleotide numbering
FT                         given in the specification for this DNA sequence"
FT   misc_feature    5881. .5940
FT                   /*tag= d
FT                   /note= "these bases represent a line of missing text in
FT                         the sequence listing in the specification. They
FT                         are included to maintain the nucleotide numbering
FT                         given in the specification for this DNA sequence"
FT   misc_feature    7681. .7740
FT                   /*tag= e
FT                   /note= "these bases represent a line of missing text in
FT                         the sequence listing in the specification. They
FT                         are included to maintain the nucleotide numbering
FT                         given in the specification for this DNA sequence"
FT   misc_feature    9481. .9540
FT                   /*tag= f
FT                   /note= "these bases represent a line of missing text in
FT                         the sequence listing in the specification. They
FT                         are included to maintain the nucleotide numbering
FT                         given in the specification for this DNA sequence"
FT   misc_feature    11281. .11340
FT                   /*tag= g
FT                   /note= "these bases represent a line of missing text in
FT                         the sequence listing in the specification. They
FT                         are included to maintain the nucleotide numbering
FT                         given in the specification for this DNA sequence"
FT   misc_feature    13081. .13140
FT                   /*tag= h
FT                   /note= "these bases represent a line of missing text in
FT                         the sequence listing in the specification. They
FT                         are included to maintain the nucleotide numbering
FT                         given in the specification for this DNA sequence"
FT   misc_feature    14881. .14940
FT                   /*tag= i
FT                   /note= "these bases represent a line of missing text in
FT                         the sequence listing in the specification. They
FT                         are included to maintain the nucleotide numbering
FT                         given in the specification for this DNA sequence"
PN   EP-786519-A2.
PD   30-JUL-1997.
PF   07-JAN-1997; 100117.
PR   05-JAN-1996; US-009861.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI   Rosen CA;
DR   WPI; 97-374922/35.
PT   Polynucleotide(s) and proteins derived from Staphylococcus aureus
PT   stored on computer readable medium and used in the production of
PT   anti-S.aureus vaccines
PS   Claim 1; Page 531-540; 3271pp; English.
CC   This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC   of the invention. The DNA sequences are recorded on a computer readable
CC   medium, preferably selected from a floppy or hard disk, random access
CC   memory (RAM), read only memory (ROM) or CD ROM. Homology searches using
CC   the S.aureus DNA sequences allows putative functions to be assigned so
CC   that protein encoding or regulatory regions of commercial, therapeutic or
CC   industrial importance can be obtained. Specifically, sequences which are
CC   likely to encode antigens have been identified and these polypeptides can
CC   be used in a vaccine composition against S.aureus infection. The
CC   polypeptides can also be used in a kit for the immunodetection of
CC   S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC   including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC   skin and surgical wound infections, scalded skin syndrome, toxic shock
CC   syndrome, etc. Organisms transformed with the DNA sequences can be used
CC   for recombinant production of the polypeptides. The new DNA sequences
CC   (and their fragments) are useful as primers or probes for isolating
CC   homologues of any of the S.aureus DNA sequences contained on the
CC   computer readable medium.
SQ   Sequence   15598 BP;    5132 A;    2487 C;    2962 G;    4470 T;


  Query Match             3.4%;  Score 35.4;  DB 1;  Length 15598;
  Best Local Similarity   61.3%;  Pred. No. 2;
  Matches   57;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy      94 gttctacacaaattaacaattcgtccaaaaagaggtatttaagcgaagatttggcattttt 153
           | || |  | ||||| ||||| || |  |||| | |||  |   ||   ||| | |  ||
Db    9714 GCTCCAATCCAATTACCAATTGGTGCGGAAGACGAATTCGAAGCAATCATTGACTTAGTT 9773

Qy     154 ttaatccatttttaaaaaaatacatctgcttta 186
             |||  | | ||  ||| |||||  || ||||
Db    9774 GAAATGAAATGTTTCAAATATACAAATGATTTA 9806


RESULT  13
X20248_00/c
WP Sequence split into 10 fragments  locus X20248 Accession X20248
WP      Fragment Name      Begin      End
WP      X20248_00              1    110000
WP      X20248_01         100001    210000
WP      X20248_02         200001    310000
WP      X20248_03         300001    410000
WP      X20248_04         400001    510000
```

```
WP      X20248_05          500001    610000
WP      X20248_06          600001    710000
WP      X20248_07          700001    810000
WP      X20248_08          800001    910000
WP      X20248_09          900001    910715
ID   X20248 standard; DNA; 910715 BP.
AC   X20248;
DT   04-MAY-1999  (first entry)
DE   Borrelia burgdorferi polynucleotide sequence #1.
KW   Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
KW   epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis,
KW   infection; diagnosis; characterisation; detection; ds.
OS   Borrelia burgdorferi.
PN   WO9858943-A1.
PD   30-DEC-1998.
PF   18-JUN-1998; 98WO-012764.
PR   03-SEP-1997; 97US-057483.
PR   20-JUN-1997; 97US-050459.
PR   22-JUL-1997; 97US-053344.
PR   22-JUL-1997; 97US-053477.
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (MEDI-) MEDIMMUNE INC.
PI   Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO,
PI   White OR;
DR   WPI; 99-081217/07.
PT   New isolated Borrelia burgdorferi nucleic acids  used to develop
PT   products for the detection, diagnosis, characterisation  prevention
PT   and therapy of infections, particularly Lyme disease
PS   Claim 1; Page 157-671; 1128pp; English.
CC   X20248 to X20402 represent polynucleotide sequences isolated from
CC   Borrelia burgdorferi (Bb). Products derived from Bb can be used for
CC   the detection, diagnosis, characterisation, prevention and therapy of
CC   Bb infections, e.g. Lyme disease. They can also be used for the
CC   production of biosynthetic products, e.g. enzymes. Borrelia belongs
CC   to a family of motile, spiral shaped bacteria called Spirochetes.
CC   Spirochetes are pathogenic in humans and Borrelia causes epidemic and
CC   endemic relapsing fever, and Lyme borreliosis, more commonly known as
CC   Lyme disease.
SQ   Sequence   910715 BP;    327171 A;    129646 C;    130753 G;    323091 T;


  Query Match              3.4%;  Score 35.2;  DB 1;  Length 110000;
  Best Local Similarity   52.0%;  Pred. No. 5;
  Matches   79;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy       2 atctgtaaattacatatatttatttgcgcacattttgctaaccgcctattcttttttaaaa 61
           |||     || |  ||| |||| ||||         ||| ||  ||  ||  |  |||| ||
Db    5856 ATCATGTAAATTCATAAATTTTTTTTTTCATATTTTCTCTCCTACTACAATCTTTATAA 5797

Qy      62 ttttaacgattaagaacgcaatttcaatttgcgttctacacaaattaacaattcgtccaa 121
            || |  | |    || |||| |  ||||    ||||| | ||   ||   |||
Db    5796 ATTCATAGTTGTTTAATGCAAATATATTTCTATTTCTATAAAATACAATTTTTCCATATC 5737

Qy     122 aagaggtatttaagcgaagatttgacatttt 153
             |  | |||     ||  |||   ||||||
Db    5736 GTTAATTTTTTGTTTTAATTTTTTATATTTTT 5705


RESULT  14
I78912
ID   I78912 standard; cDNA; 4200 BP.
AC   I78912;
DT   09-FEB-1998  (first entry)
DE   Spinocerebellar ataxia gene SCA2.
KW   Monoclonal antibody; neurodegenerative disease; polyglutamine; TBP;
KW   repeat region; affinity; TATA binding protein; Kennedy disease,
KW   transcription initiation factor; lymphoblastic cell line; schizophrenia;
KW   Huntington's disease; dominant autosomal spinocerebellar ataxia;
KW   X-linked spino-bulbular muscular atrophy; familial spastic paraplegia;
KW   dentatorubral-pallidoluisial atrophy; bipolar affective disorder,
KW   manic depressive psychosis; ss
OS   Homo sapiens.
FH   Key             Location/Qualifiers

FT   CDS             1. .2747
FT                   /*tag= a
FT                   /product= "SCA2 protein
FT                   /note= "this DS contains a putati
FT                   code for the SCA2 protein
FT   CDS             2594. .3640
FT                   /*tag= b
FT                   /note= "this second open reading f
FT                   by a frameshift or by alter
FT   CDS             3. .2412
FT                   /*tag= c
FT                   /note= "putative open reading fram
FT                   with the putative translat
FT                   the SCA2 open reading fram
FT   misc_signal     239. .245
FT                   /*tag= d
FT                   /note= "putative Kozak consensus s
FT   repeat_region   258. .323
FT                   /*tag= e
FT                   /note= "encodes polyglutamine repe
FT                   repeats of CAG with 2 CAA
FT   repeat_unit     258. .260
FT                   /*tag= f
FT                   /note= "CAA repeats"
FT   misc_feature    1. .998
FT                   /*tag= g
FT                   /note= "Sequence contained in CAN.
FT   misc_feature    3987. .4203
FT                   /*tag= h
FT                   /note= "derived from the EST's Bg.
FT                   from dbEST database"
FT   misc_feature    4023. .4029
FT                   /*tag= i
FT                   /note= "region which differs in 2
FT                   sequences of the EST clone
FT                   and z15574"
PN   WO9717445-A1.
PD   15-MAY-1997.
PF   08-NOV-1996; 96FR-013774.
PR   10-NOV-1995; FR-012574.
PA   (CNRS ) CNRS CENT NAT RECH SCI.
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PI   Lutz Y, Mandel J, Trottier Y;
DR   WPI; 97-281034/25.
DR   P-PSDB; W24800, W24801.
PT   Antibody 1C2 used for treating or preventing neur
PT   diseases - associated with proteins containing l
PT   repeats, e.g. Huntington's disease
PS   Claim 21; Page 45-47; 69pp; French.
CC   The invention relates to a monoclonal antibody (M
CC   treatment of neurodegenerative diseases associate
CC   of polyglutamine repeat regions.  This MAb is aff
CC   affinity to the TATA binding protein (TBP) transc
CC   factor, especially at the amino acid sequence LEF
CC   the N-terminus of TBP.  MAb 1C2 has been shown t
CC   for polyglutamine repeats with a proportional aff
CC   of glutamine repeats.  This affinity has been use
CC   encoding proteins containing long polyglutamine r
CC   implicated in neurodegenerative diseases.  A scre
CC   library, generated from a lymphoblastic cell line
CC   suffering from spinocerebellar ataxia (SCA), wit
CC   new sequences (I78906-I78911) encoding polyglutar
CC   also isolated the complete SCA2 gene in clone DAN
CC   here).  The sequence appears to contain 2 open re
CC   second of which may be generated by an frameshift
CC   alternative splicing event.  The first ORF also c
CC   polyglutamine repeat region near the N-terminus
CC   SCA2 alleles contain 17-29 CAG triplet repeats w
CC   interspersed whereas the mutant sequence from pa
CC   contains at least 30, preferably 37-50 CAG repeat
CC   MAb 1C2, active fragment of it or nucleic acids
CC   specifically used to treat Huntington's disease
CC   X-linked spino-bulbular muscular atrophy (Kennedy
```

```
CC   dentatorubral-pallidoluisial atrophy, dominant autosomal spinocerebellar
CC   ataxia, familial spastic paraplegia, bipolar affective disorder, manic
CC   depressive psychoses and schizophrenia.
SQ   Sequence   4200 BP;    1152 A;    1200 C;    913 G;    935 T;


  Query Match              3.2%;  Score 33.2;  DB 1;  Length 4200;
  Best Local Similarity   48.9%;  Pred. No. 5;
  Matches   89;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy     193 ctttgccaaggaatacc-gactagcataacctbgattacaaaaaacggctaaaagctgtcgg 252
           || |  ||  |   | | | || |  |||  || || || |||  ||       |
Db    3864 CTGTTTCAGAGTCCCCCAGGTACCCCAGCTCTGCTTGCCCAAACTGGAAGTTATTTATTT 3923

Qy     253 tctatactatattaaaaatctaaaaaactaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaat 312
           | ||     | ||    |   ||||||  | ||| |     || ||      |  | | |
Db    3924 TTTAATAACCCTTGAAAGTCATGAACACATCAGCTAGCAAAAGAAGTAACAAGAGTGATT 3983

Qy     313 catgctgctagtaaatgcgatataaatgggccaaaacaaacgtactaqaaatgactaaataagtac 372
           | |||||||| || ||| | | ||     ||| ||   |  | |||| |  | |  | ||
Db    3984 CTTGCTGCTATTACTGCTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATCAAGACTTGGAAC 4043

Qy     373 cc 374
           |
Db    4044 GC 4045


RESULT  15
V21209_04/c
Continuation (5 of 17) of V21209 from base 400001 (Methanococcus jannaschii circular chr
WP Sequence split into 17 fragments  LOCUS V21209 Accession V21209
WP      Fragment Name      Begin      End
WP      V21209_00              1     110000
WP      V21209_01         100001     210000
WP      V21209_02         200001     310000
WP      V21209_03         300001     410000
WP      V21209_04         400001     510000
WP      V21209_05         500001     610000
WP      V21209_06         600001     710000
WP      V21209_07         700001     810000
WP      V21209_08         800001     910000
WP      V21209_09         900001    1010000
WP      V21209_10        1000001    1110000
WP      V21209_11        1100001    1210000
WP      V21209_12        1200001    1310000
WP      V21209_13        1300001    1410000
WP      V21209_14        1400001    1510000
WP      V21209_15        1500001    1610000
WP      V21209_16        1600001    1664976


  Query Match              3.1%;  Score 32.8;  DB 1;  Length 110000;
  Best Local Similarity   58.0%;  Pred. No. 25;
  Matches   58;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy      76 aacgcaattcaatttgcgttctacacaaattaacaattcgtccaaaaqaqgtatttaaq 135
           || || |||  |   |     |  |   | || |||||     |||| |   |||||| |
Db 106947 AAGGCTATTGGAAGATTGCATAAATTAATATCAACAAAAAACCAAGAATTGTATTTGAA 106888

Qy     136 cgaagatttgucatttttttaatccattttaaaaaaata 175
           || || ||||| |||| |||  |||| | | |||  | |
Db 106887 AGAGCAATTGGCTTTTTATTATACCATATATTAAATGAGA 106848


Search completed: August 25, 2000, 22:57:07
Job time: 9463 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd

OM nucleic - nucleic search, using sw model

Run on: August 25, 2000, 19:29:03 , Search time 43.31 Seconds
(without alignments)
3425.451 Million cell updates/sec

Title: US-09-257-585-1
Perfect score: 1047
Sequence: 1 gatctgtaaattacatatat ... atgaggcgacggcggatcc 1047

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 243088 seqs, 68777415 residues

Total number of hits satisfying chosen parameters: 486160

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/6_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	37.6	3.6	2058	4	US-08-749-391-1	Sequence 1, Appli
2	35.8	3.4	7218	1	US-08-232-463-14	Sequence 14, Appl
3	32.8	3.1	19877	4	US-08-816-155B-8	Sequence 8, Appli
c 4	32	3.1	721	2	US-08-451-405A-2	Sequence 2, Appli
5	31.4	3.0	858	1	US-08-126-543A-1	Sequence 1, Appli
6	31.4	3.0	858	2	US-08-454-039A-1	Sequence 1, Appli
c 7	31.2	3.0	1422	2	US-08-319-704-5	Sequence 5, Appli
8	31.2	3.0	2502	2	US-08-062-472B-2	Sequence 2, Appli
c 9	31.2	3.0	5852	1	US-07-867-106-2	Sequence 2, Appli
c 10	30.8	2.9	3865	2	US-08-832-883-48	Sequence 48, Appl
c 11	30.8	2.9	3865	3	US-08-832-877-48	Sequence 48, Appl
c 12	30.6	2.9	4507	3	US-08-568-459A-3	Sequence 3, Appli
c 13	30.6	2.9	4507	4	US-08-487-826B-3	Sequence 3, Appli
c 14	30.4	2.9	4380	1	US-07-582-945-1	Sequence 1, Appli
c 15	30.4	2.9	4380	3	US-08-453-141-1	Sequence 1, Appli
16	30.2	2.9	1262	1	US-07-715-751B-3	Sequence 3, Appli
17	30.2	2.9	1939	1	US-07-715-751B-2	Sequence 2, Appli
18	30.2	2.9	2215	4	US-08-980-329C-1	Sequence 1, Appli
c 19	30.2	2.9	4566	2	US-08-484-101B-37	Sequence 37, Appl
20	30	2.9	1601	1	US-07-715-751B-1	Sequence 1, Appli
c 21	30	2.9	6973	2	US-08-478-370-1	Sequence 1, Appli
c 22	29.8	2.8	471	1	US-08-112-816-1	Sequence 1, Appli
c 23	29.8	2.8	8041	2	US-08-765-081-1	Sequence 1, Appli
c 24	29.8	2.8	8041	5	US-09-098-082-1	Sequence 1, Appli
c 25	29.8	2.8	8041	6	PCT-US95-06994-1	Sequence 1, Appli
26	29.6	2.8	2523	4	US-08-410-784A-3	Sequence 3, Appli
c 27	29.6	2.8	3193	4	US-08-868-786-1	Sequence 1, Appli
28	29.4	2.8	3178	5	US-08-968-564-7	Sequence 7, Appli
c 29	29.2	2.8	644	1	US-08-129-151A-6	Sequence 6, Appli
c 30	29.2	2.8	644	4	US-08-715-579-6	Sequence 6, Appli
31	29.2	2.8	1415	1	US-08-413-118-127	Sequence 127, Ap
32	29.2	2.8	1415	5	US-08-473-446-127	Sequence 127, Ap
c 33	29.2	2.8	2529	4	US-08-410-784A-1	Sequence 1, Appli
34	29.2	2.8	3738	3	US-08-480-473B-1	Sequence 1, Appli
35	29.2	2.8	3738	5	US-08-915-213-1	Sequence 1, Appli
36	29.2	2.8	3738	6	PCT-US96-10251-1	Sequence 1, Appli
37	29.2	2.8	5852	1	US-07-867-106-2	Sequence 2, Appli
c 38	29.2	2.8	19124	4	US-08-487-826B-13	Sequence 13, Ap
39	29	2.8	1544	3	US-08-867-057-2	Sequence 2, Appl
40	29	2.8	1544	4	US-09-128-369-2	Sequence 2, Appl
c 41	29	2.8	4526	1	US-07-855-412B-4	Sequence 4, Appl
c 42	29	2.8	4526	3	US-08-308-887A-4	Sequence 4, Appl
c 43	29	2.8	4526	5	US-08-881-094-4	Sequence 4, Appl
c 44	29	2.8	9837	2	US-08-832-883-68	Sequence 68, Ap
c 45	29	2.8	9837	3	US-08-832-877-68	Sequence 68, Ap

ALIGNMENTS

RESULT 1
US-08-749-391-1/c
; Sequence 1, Application: US/08749391
; Patent No. 5948667
; GENERAL INFORMATION:
; APPLICANT: Cheng, Kuo-Joan
; APPLICANT: Schinger, Leonard B.
; APPLICANT: Liu, Jin Hao
; APPLICANT: Hu, Youji
; APPLICANT: Forsberg, Cecil W.
; APPLICANT: Mobley, Maurice M.
; TITLE OF INVENTION: A Xylanase Obtained From an
; TITLE OF INVENTION: Anaerobic Fungus
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.3
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/749,391
; FILING DATE: 13-NOV-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Donna M. Ferbo
; REGISTRATION NUMBER: 34,878
; REFERENCE/DOCKET NUMBER: 94-96
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2058 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Neocallimastix patriciarum
; STRAIN: 27

```
;     IMMEDIATE SOURCE:
;       LIBRARY:  genomic DNA library
;       CLONE:  pNspX-06
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  301..1755
US-08-749-391-1


  Query Match              3.6%;  Score 37.6;  DB 4;  Length 2058;
  Best Local Similarity   56.5%;  Pred. No. 0.056;
  Matches   78;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy       2 a*e*g*aa**t***a*a*atatttat***tgcgcgtattttjstaaccgcctattttttttaaaa 61
           || | |   ||  |||||||| ||||     |||||  |||     |||  ||| |||||
Db    2031 ATATTTTTTTTTTATATATTTTTTTTTATAATTTTATTAAAATAGTATCTTTTATAAAA 1972

Qy      62 ****acgattaaaaaacgcattgaaattttgcgttctacacaaattaacaattcgtccaa 121
           | ||||  |||||  |     || || ||| | ||        |  ||  ||||    ||
Db    1971 TATTAAATATTAAATAAATTGTTTAAATTTACATTTACATTTACGTATTAATTAAAAAAA 1912

Qy     122 aaqa 125
           ||||
Db    1911 AAAA 1908


RESULT   2
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22314-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,213
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  50472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pIZgpt-Fls
```

```
US-08-232-463-14


  Query Match              3.4%;  Score 35.8;  DB 1;  Length 7218;
  Best Local Similarity   54.7%;  Pred. No. 0.35;
  Matches   46;  Conservative  136;  Mismatches  153;  Indels    0;  Gaps    0;

Qy     386 cttgctgtaqtattatatttaacgcaagcgtccatttactgccagqqcaagtttatcaaatq 445
           ::: :: :  : :: : ::: :   :  : ::: ::: :: ::    :   ::: :: :
Db    1203 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1262

Qy     446 gqaccaqtgttctccctcctctqqacaactcagttctttqcaaacgctaqacagtctacc 505
           ::  : ::::::::::::::  :  ::: :::::: :   : ::   :  ::: ::
Db    1263 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1322

Qy     506 tctctqccacccatttttacttttcaaaaatttactcttqccqctactqaaacttctaca 565
           ::::: :: :: ::::: ::::::     ::: :::::: :: :: ::    ::::: :
Db    1323 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1382

Qy     566 attgaaagagccccacaatqaaaqtccaaqctacattqcgcaccccttatccqcccttqccqqc 625
           ::       ::: :  :     :::  :: : ::: :: ::::: :: :: |  |
Db    1383 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYGTACCAAA 1442

Qy     626 ttactttccagccagccaatqcttqqaaaqattqcatcatccaacgttataaaqacqqccqa 685
           ||  | |    |    | | |||| | |||| |    | ||  | |    | | | ||
Db    1443 TTCTTCTATCTCTTTAACTACTTGCATAGATAGCTAATTACAGTGATGCCTACATGCCGT 1502

Qy     686 tqtcaacaacatatatactqccaataqqaacqaaq 720
           | |   |||  |    |||  ||| | ||| |
Db    1503 TTTTTGAAACTGAATAGATGCGTCTAGAACCCATG 1537


RESULT   3
US-08-816-155B-8
; Sequence 8, Application US/08816155B
; Patent No. 5990091
;  GENERAL INFORMATION:
;    APPLICANT:  TARTAGLIA, JAMES
;    APPLICANT:  COX, WILLIAM I.
;    APPLICANT:  GETTIG, RUSSELL R.
;    APPLICANT:  MARTINEZ, HECTOR
;    APPLICANT:  PAOLETTI, ENZO
;    APPLICANT:  PINCUS, STEVEN E.
;    TITLE OF INVENTION:  VECTORS HAVING ENHANCED EXPRESSION, AND
;    TITLE OF INVENTION:  METHODS OF MAKING AND USES THEREOF
;    NUMBER OF SEQUENCES:  48
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FROMMER LAWRENCE & HAUG LLP
;      STREET:  745 FIFTH AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10151
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/816,155B
;      FILING DATE:  12-MAR-1997
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOWALSKI, THOMAS J.
;      REGISTRATION NUMBER:  32,147
;      REFERENCE/DOCKET NUMBER:  454310-2990
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 588-0800
;      TELEFAX:  212-588-0500
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  19877 base pairs
;      TYPE:  nucleic acid
```

```
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-08-816-155B-8


  Query Match              3.1%;  Score 32.8;  DB 4;  Length 19877;
  Best Local Similarity   56.5%;  Pred. No. 4.5;
  Matches   61;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy      15 atatatttatttqccqcatttqctaaccqcctattcttttaaaatttaacqattaa 74
           ||||| |||||||||    ||  || ||||     |  || ||| | ||  |||   |||||
Db    2148 AIATAATTAITTIGIAAGAIGAIGTIAACTATGTCAICGTTICATAACATTATGTATTAA 2207

Qy      75 qaacqcaattcaatttqcqttctacacaaattaacaattcqtccaaa 122
           |||  |||||        ||| |   | | | || | |    || | |
Db    2208 GAAGACAATTATGCAGAACGTGCATTAGATAATATCTACGGTTCTATA 2255


RESULT   4
US-08-451-405A-2/c
; Sequence 2, Application US/08451405A
; Patent No. 5736358
;  GENERAL INFORMATION:
;    APPLICANT:  FASEL, NICOLAS JOSEPH
;    APPLICANT:  REYMOND, CHRISTOPHE DOMINIQUE
;    TITLE OF INVENTION:  DICTYOSTELID EXPRESSION VECTOR AND
;    TITLE OF INVENTION:  METHOD FOR EXPRESSING A DESIRED PROTEIN
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THE WEBB LAW FIRM
;      STREET:  700 KOPPERS BUILDING, 436 SEVENTH AVENUE
;      CITY:  PITTSBURGH
;      STATE:  PENNSYLVANIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  15219-1818
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" FLOPPY DISK
;      COMPUTER:  Midwest Micro 486-50
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  WORDPERFECT 6.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/451,405A
;      FILING DATE:  26-MAY-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/965,273
;      FILING DATE:  15-JAN-1993
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  731
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  SINGLE
;      TOPOLOGY:  UNKNOWN
US-08-451-405A-2


  Query Match              3.1%;  Score 32;  DB 2;  Length 731;
  Best Local Similarity   51.0%;  Pred. No. 1.7;
  Matches  102;  Conservative    0;  Mismatches   95;  Indels    3;  Gaps    1;

Qy       2 atctgtaaattacatatatttatttgccgcattttgctaaccgcctattcttttaaaa 61
           |||   |||||| |  |||| ||||   | ||||| ||    | ||| |  |   |||
Db     252 ATCCAAAAATTATCTGAATTTTTTTTTTAGAATTTTCTTA---TCATATACCGTCACAAA 196

Qy      62 ttttaacgattaagaacgcaattcaatttqcgttctacacaaattaacaattcgtccaa 121
           | |          ||    || ||| |    ||  | |||||| || | | | | || ||
Db     195 TCTATTTTTAGGTTCACTATGTTTAATATAATTTTAAATACAAATAAAAACTCCTTCAAA 136

Qy     122 aagaggtatttaagcgaagattgqcattttttaatccattttaaaaaaatacatctq 181
           |||  ||||| |    :  |||    ||||   ||   |    ||||| | | |   |
Db     135 TAGAAATATTTTATTCTAATTTTTATTTTTAAAAAAATAAAAATAAAAATAAAAAAAG 76

Qy     182 ctttaacccacctttqccaa 201
           ||   || | | | ||
Db      75 GAAAAAAAAACATATATTAA 56


RESULT   5
US-08-126-593A-1
; Sequence 1, Application US/08126593A
; Patent No. 5527700
;  GENERAL INFORMATION:
;    APPLICANT:  Kaslow, David C.
;    APPLICANT:  Duffy, Patrick E.
;    TITLE OF INVENTION:  Target Antigens of Transmissi
;    TITLE OF INVENTION:  Antibodies for Malaria Parasi
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew
;      STREET:  One Market Plaza, Steuart Street Tower
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94105-1492
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/126,593A
;      FILING DATE:  22-SEP-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US  7/91..294
;      FILING DATE:  10-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bastian, Kevin L.
;      REGISTRATION NUMBER:  34,774
;      REFERENCE/DOCKET NUMBER:  15280.46-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 543-9600
;      TELEFAX:  (415) 543-5043
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  858 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  123..79
;      OTHER INFORMATION:  /product= "Pqs28 surface pr
;      OTHER INFORMATION:  /note= "from Plasmodium gallina
US-08-126-593A-1


  Query Match              3.1%;  Score 31.4;  DB 1;
  Best Local Similarity   52.7%;  Pred. No. 2.8;
  Matches   68;  Conservative    0;  Mismatches   61;  Indels

Qy      50 ttttttttaaaatttttaacqattaaaaacqcaattcaatttqc
           || || | | | | | |  |  |  | | || ||| :::
Db       2 TTTTTGTCATATTATTATCATTTTTAAATTCATTTCTATTTCCC

Qy     110 caattcqtccaaaaqaqqatatttaaacqaaqatttqqcattttt
           |||    | |||| | |  | :  :      ||       |: :::
Db      62 CAAAATATTCAAACCAATATTATTTATGTAAACCAAAACAATTTT

Qy     170 aaaataqat 178
           ||   | ||
Db     122 AATGAAAAT 130
```

```
RESULT   6
US-08-454-039A-1
; Sequence 1, Application US/08454039A
; Patent No. 5753238
;  GENERAL INFORMATION:
;    APPLICANT:  Kaslow, David C.
;    APPLICANT:  Duffy, Patrick E.
;    TITLE OF INVENTION:  Target Antigens of Transmission Blocking
;    TITLE OF INVENTION:  Antibodies for Malaria Parasites
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew LLP
;      STREET:  Two Embarcadero Center, Eighth Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/454,039A
;      FILING DATE:  30 MAY-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/912,294
;      FILING DATE:  10-JUL-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/126,593
;      FILING DATE:  22-SEP-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bastian, Kevin I
;      REGISTRATION NUMBER:  34,774
;      REFERENCE/DOCKET NUMBER:  15280 46-1-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 576 0200
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  858 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Plasmodium gallinaceum
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  123..791
;      OTHER INFORMATION:  /product= "Pqs28"
US-08-454-039A-1


  Query Match             3.0%;  Score 31.4;  DB 2;  Length 858;
  Best Local Similarity   52.7%;  Pred. No. 2.8;
  Matches   68;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy      50 ttcttttttaaaatttttaacgattaaaaacgcaattcaattttgcgttctacacaaaattaa 109
           || || | | | | ||| | ||  ||  || ||| |||| | |  || | | | |  |
Db       2 TTTTTGTCATATTATTATCATTTTAAATTCATTTCTATTTCCCATAATAAATTATTCTA 61

Qy     110 caattcgtccaaaaqaaqataattaaacgaaaqattqqcqattttttaatccattttaaa 169
           |||    | |||| || |  | |     ||       || ||||   ||  ||||  |||
Db      62 CAAAATATTCAAACGAAGATTATTTAGTAAACCAAAACAATTTTAACATTTATTTAAAAA 121

Qy     170 aaaatacat 178
           ||   | ||
Db     122 AATGAAAAT 130


RESULT   7
```

```
US-08-319-704-5/c
; Sequence 5, Application US/08319704
; Patent No. 5814617
;  GENERAL INFORMATION:
;    APPLICANT:  Hoffman, Stephen L.
;    APPLICANT:  Charoenvit, Yupin
;    APPLICANT:  Hedstrom, Richard C.
;    APPLICANT:  Doolan, Denise L.
;    TITLE OF INVENTION:  Protective 17 kDa Malaria Hepatic and
;    TITLE OF INVENTION:  Erythrocytic Stage Immunogen and Gene
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Naval Medical R & D Command
;      STREET:  Bldg 1, T-12, 8901 Wisconsin Avenue
;      CITY:  Bethesda
;      STATE:  Maryland
;      COUNTRY:  U.S.A
;      ZIP:  20889-5606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/319,704
;      FILING DATE:  07-OCT-1994
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  A. David Spevack
;      REGISTRATION NUMBER:  24,743
;      REFERENCE/DOCKET NUMBER:  75,206
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 295-6759
;      TELEFAX:  (301) 295-1022
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1422 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-319-704-5


  Query Match             3.0%;  Score 31.2;  DB 2;  Length 1422;
  Best Local Similarity   48.8%;  Pred. No. 4.1;
  Matches   84;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy       7 taaattacatatatttattttgccgcattttgctaaccgcctattcttttttaaaatttta 66
           |||| || | |||| || ||||  | |  |    ||      |  |  | ||  | | | |
Db     357 TAAAATAAAGATATATACTTTGAAGAAAATCTTAAATTATATGCTTATATTGGATTATAA 298

Qy      67 acgattaaqaacgcaattcaattttgcgttctacacaaattaacaattcgtccaaaaqaq 126
           || | |  |    |     ||     || | |  | ||| | ||  |  ||||
Db     297 TTGAATTAATAAAATAAATTTTTAATTATTATTGAATTAATAATATTTTATTAAAAATTT 238

Qy     127 qtatttaaqcgaagatttgqcatttttttaatccatttttaaaaaaatacat 178
           |||  ||  ||   |    ||| |||| || ||   |||| || | | ||
Db     237 AAATTCTAGAAAAATATAAATATTATTTTCATACAAAATTAATAATAAAAAT 186


RESULT   8
US-08-062-472B-2
; Sequence 2, Application US/08062472B
; Patent No. 5695954
;  GENERAL INFORMATION:
;    APPLICANT:  Sherwood, Nancy G M
;    APPLICANT:  Parker, David B
;    APPLICANT:  McRory, John E
;    APPLICANT:  Lescheid, David W
;    TITLE OF INVENTION:  DNA ENCODING TWO FISH NEUROPEPTIDES
;    NUMBER OF SEQUENCES:  49
;    CORRESPONDENCE ADDRESS:
```

```
;       ADDRESSEE:  KLARQUIST, SPARKMAN, CAMPBELL, LEIGH &
;       ADDRESSEE:  WHINSTON, LLP
;       STREET:  ONE WORLD TRADE CENTER, SUITE 1600, 121 S.W.
;       STREET:  SALMON STREET
;       CITY:  PORTLAND
;       STATE:  OREGON
;       COUNTRY:  USA
;       ZIP:  97204-2988
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/062,472B
;       FILING DATE:  14-MAY-1993
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  POLLEY, RICHARD J
;       REGISTRATION NUMBER:  28107
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (503) 226-7391
;       TELEFAX:  (503) 228-9446
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2502 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-062-472B-2


  Query Match                3.0%;  Score 31.2;  DB 2;  Length 2502;
  Best Local Similarity   57.0%;  Pred. No. 5.3;
  Matches   57;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy      96 tctacacaaaattaacaattcgtccaaaaagaggtatttaagcgaagatttggcattttttt 155
           | | ||||   ||      ||||||  |  |||| | |      |   |||||||
Db    1759 TGTGCACAATCAAAGGTCAGGTCCAATTGTTTATTAAGTCCCCCCATATATATTTTTTA 1818

Qy     156 aatccatttttaaaaaaatacatctgtttaacccacctt 195
           ||||| | ||||||||||| |  | | ||| | ||   ||
Db    1819 AATCCTTATTTAAAAAAACATTCCAGGTTTCAGCCTATTT 1858


RESULT   9
US-07-867-106-2/c
; Sequence 2, Application US/07867106
; Patent No. 5389526
;  GENERAL INFORMATION:
;    APPLICANT:  Slade, Martin B
;    APPLICANT:  Chang, Andy C M
;    APPLICANT:  Williams, Keith L
;    TITLE OF INVENTION:  Improved Plasmid Vectors for Cellular
;    TITLE OF INVENTION:  Slime Moulds of the Genus Dictyostelium
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris
;      STREET:  One Liberty Place 46th Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/867,106
;      FILING DATE:  19920625
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  AU PI 7183
;      APPLICATION NUMBER:  PCT/AU90/00530
;      FILING DATE:  02-NOV-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Feeney, Joanne Lucas
;      REGISTRATION NUMBER:  35,744
;      REFERENCE/DOCKET NUMBER:  5389-0002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-568-3100
;      TELEFAX:  215-568-3439
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5852 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2378..5038
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2378..5038
US-07-867-106-2


  Query Match                3.0%;  Score 31.2;  DB 1;  Length
  Best Local Similarity   48.8%;  Pred. No. 7.8;
  Matches   84;  Conservative    0;  Mismatches   88;  Indels

Qy      16 tatatttatttttgccaattttgctaaccgctattcttttaaa... 73
           | | ||  ||||| || | |||||          | | ||||
Db     493 TTTTTTCATTTTCCCACTTTTTACCATTTTTTTCTTTTTTTC... 544

Qy      76 aacgcaattcaatttaccattcatcaaaattaaccaattcgtcca... 135
           ||   || | ||| |   |  | |  ||||||  | |  |  |
Db     433 AAGTTAAATAAATAATCAAATATTCAACAATTAAATAATTATTTA... 487

Qy     136 cgaaqatttqqcattttttaatccatttttaaaaaaatacatct... 195
           | | |||   | ||||  |  | ||| ||||
Db     373 CTCATTTTTTATAAATTTACATTTATAAAAAATTTTTA... 422


RESULT  10
US-08-832-883-48/c
; Sequence 48, Application US/08832883
; Patent No. 5807681
;  GENERAL INFORMATION:
;    APPLICANT:  Giordano, Antonio
;    APPLICANT:  Baldi, Alfonso
;    TITLE OF INVENTION:  METHODS FOR THE DIAGNOSIS AND
;    TITLE OF INVENTION:  OF CANCER
;    NUMBER OF SEQUENCES:  115
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEIDEL, GONDA, LAVORGNA & MONACO, P.
;      STREET:  Suite 1800 Two Penn Center Plaza
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/832,883
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Monaco, Daniel A
;      REGISTRATION NUMBER:  30,480
```

; REFERENCE/DOCKET NUMBER: 8321-13 US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-8383
; TELEFAX: (215) 568-5549
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3865 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-832-883-48

Query Match 2.9%; Score 30.8; DB 2; Length 3865;
Best Local Similarity 45.7%; Pred. No. 8.6;
Matches 107; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

Qy 387 ttgctgtaqtattatatttaacgaaqcgtccatttactgccagggcaaqtttatcaatqg 446
Db 969 TTGTCATTCCATTTTATTTTTAAAAGGTTACACTTAATGTAAACATCCTTCACAGTCTGG 910

Qy 447 gaccagtgttctccctcctctggacaactcagttctttgcaaacgctagacagtctacct 506
Db 909 GAAAACTTTTATATTTAAAGAGAAGCAAACATTATTTATTCTAGCCAAATAATTTTAACT 850

Qy 507 ctctgccaccattttttacttttcaaaaattttactcctttgccgctactgaaacttctacaa 566
Db 849 CCTTTCCCTTTACTTTTCTTGTCAACATTTTTTTAATTGTTAGAATTTTCATTTCTACTT 790

Qy 567 ttgaaaqaqcccacaatgaaaqtccaaqctacattgccaccctttatgcccctt 620
Db 789 TATGAAAAAACTTTCAATGATAAAGAAGGAAAATGTGTATTCCATAACTGCCTT 736

RESULT 11
US-08-832-877-48/c
; Sequence 48, Application US/08832877
; Patent No. 5840506
; GENERAL INFORMATION:
; APPLICANT: Giordano, Antonio
; TITLE OF INVENTION: METHODS FOR THE DIAGNOSIS AND PROGNOSIS OF
; TITLE OF INVENTION: CANCER
; NUMBER OF SEQUENCES: 116
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEIDEL, GONDA, LAVORGNA & MONACO, P.C.
; STREET: Suite 1800 Two Penn Center Plaza
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/832,877
; FILING DATE:
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: Monaco, Daniel A
; REGISTRATION NUMBER: 30,480
; REFERENCE/DOCKET NUMBER: 8321-13 US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-8383
; TELEFAX: (215) 568-5549
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3865 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-832-877-48

Query Match 2.9%; Score 30.8; DB 3; Length 3865;
Best Local Similarity 45.7%; Pred. No. 8.6;
Matches 107; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

Qy 387 ttgctgtaqtattatatttaacgaaqcgtccatttactgccagggcaaqtttatcaatqg 446
Db 969 TTGTCATTCCATTTTATTTTTAAAAGGTTACACTTAATGTAAACATCCTTCACAGTCTGG 910

Qy 447 gaccagtgttctccctcctctggacaactcagttctttgcaaacgctagacagtctacct 506
Db 909 GAAAACTTTTATATTTAAAGAGAAGCAAACATTATTTATTCTAGCCAAATAATTTTAACT 850

Qy 507 ctctgccaccattttttacttttcaaaaattttactcctttgccgctactgaaacttctacaa 566
Db 849 CCTTTCCCTTTACTTTTCTTGTCAACATTTTTTTAATTGTTAGAATTTTCATTTCTACTT 790

Qy 567 ttgaaaqaqcccacaatgaaaqtccaaqctacattgccaccctttatgcccctt 620
Db 789 TATGAAAAAACTTTCAATGATAAAGAAGGAAAATGTGTATTCCATAACTGCCTT 736

RESULT 12
US-08-568-459A-3/c
; Sequence 3, Application US/08568459A
; Patent No. 5849306
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellems, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/568,459A
; FILING DATE: 07-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4507 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
US-08-568-459A-3

Query Match 2.9%; Score 30.6; DB 3; Length 4507;
Best Local Similarity 58.1%; Pred. No. 11;
Matches 54; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Qy 3 tctgtaaattacatatatttatttttqccgcattttgctaaccgcctattctttttaaaat 62
Db 1263 TCTGAAATTTTGATTAAATAATTTTCCGCATTTTCCTTTGGAACTTTTTGAGTTTCATAT 1204

Qy 63 tttaacgattaagaacgcaattcaatttqcgt 95
Db 1203 TCTTTCGATAACGTATGCCATTCAAATTTACTT 1171

RESULT 13
US-08-487-826B-3/c
; Sequence 3, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellems, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121 001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4507 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
US-08-487-826B-3

Query Match 2.9%; Score 30.6; DB 4; Length 4507;
Best Local Similarity 58.1%; Pred. No. 11;
Matches 54; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Qy 3 tctgtaaattacatatatttatttttqccgcattttgctaaccgcctattctttttaaaat 62
Db 1263 TCTGAAATTTTGATTAAATAATTTTCCGCATTTTCCTTTGGAACTTTTTGAGTTTCATAT 1204

Qy 63 tttaacgattaagaacgcaattcaatttqcgt 95
Db 1203 TCTTTCGATAACGTATGCCATTCAAATTTACTT 1171

RESULT 14
US-07-582-945-1/c
; Sequence 1, Application US/07/582,945
; Patent No. 5369019
; GENERAL INFORMATION:
; APPLICANT: TAKKER PORED, Nicos
; APPLICANT: PETERSON, Svend
; TITLE OF INVENTION: A [illegible] VACCINE
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC DOS/MS DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/582,945
; FILING DATE: 19901012
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: [illegible]/112 PLVI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)684 4189
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4380 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Pasteurella multocida
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 219..4074
US-07-582-945-1

Query Match 2.9%; Score 30.4; DB 1; Le
Best Local Similarity 50.7%; Pred. No. 12;
Matches 73; Conservative 0; Mismatches 71; I

Qy 2 atctgtaaattacatatatttatttttqccgcattttgctaaccgcc
Db 4191 ATTGGTAATAACATTGATTTATCTATCTCATTGATTTTGATTCT

Qy 62 ttttaacgattaagaacgcaattcaatttqcgtttctaccaaaatt
Db 4131 CTTGAAAGAAAATTAAAGCCTTATAATATAAGGCTTTAATACAGTT

Qy 122 aagaggtatttaagctaaaaatt 144
Db 4071 GTGCTCTTGTTAAGCGAGCTTTT 4049

RESULT 15
US-08-453-141-1/c
; Sequence 1, Application US/08/453,141

```
; Patent No. 5885589
; GENERAL INFORMATION:
;    APPLICANT:  FOGED, Niels I.
;    APPLICANT:  PETERSEN, Svend
;    TITLE OF INVENTION:  PASTEURELLA VACCINE
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N W , Suite 500
;      CITY:  Washington, D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/453,141
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/293,314
;      FILING DATE:  22-AUG-1994
;      APPLICATION NUMBER:  US 07/582,945
;      FILING DATE:  12-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/DK89/00084
;      FILING DATE:  11-APR-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DK 1995/88
;      FILING DATE:  04-DEC-1988
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  4688.102/AKSO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4380 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  219..4076
US-08-453-141-1


  Query Match             2.9%;  Score 30.4;  DB 3;  Length 4380;
  Best Local Similarity   50.7%;  Pred. No. 12;
  Matches   73;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy       2 atctgtaaaattacatatatttatttttgccgcatttgctaaccgcctattctttttaaaa 61
           ||  ||||       |  ||||  | |  | ||||    |      | | | |||  |||
Db    4191 ATTGGTAATAACATTTCATTTTATCTATCTCATTCATTTTGATTCTTCTACTTAATAATT 4132

Qy      62 tttaaccgattaagaaccgcaattcaatttgcgttctacacaaattaacaatttcgtccaa 121
             || || ||  |  || ||  |  ||| |   | | ||    | ||   |||| ||   |
Db    4131 CTTGAAAGAAAATTAAAGCCTTATAATATAAGGCTTTAATACAGTTTTAATTACTTATA 4072

Qy     122 aagaggtatttaagcgaagatttg 145
             |   |  |||||||| | ||||
Db    4071 GTGCTCTTGTTAAGCGAGGCTTTG 4048



Search completed: August 25, 2000, 22:53:18
Job time: 12255 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2000, 17:31:26 ; Search time 579.87 Seconds
(without alignments)
7969.156 Million cell updates/sec

Title: US-09-257-585-1
Perfect score: 1047
Sequence: 1 gatctgtaaattacatatat.....atggaggcgaccgccgatcc 1047

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5247842 seqs, 2204914090 residues

Total number of hits satisfying chosen parameters: 10495684

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: qb_est1:*
21: qb_est2:*
22: qb_est3:*
23: qb_est4:*
24: qb_est5:*
25: qb_est6:*
26: qb_est7:*
27: qb_est8:*
28: qb_est9:*
29: qb_est10:*
30: qb_est11:*
31: qb_est12:*
32: qb_est13:*
33: qb_est14:*
34: qb_est15:*
35: qb_est16:*
36: qb_est17:*
37: qb_est18:*
38: qb_est19:*
39: qb_est20:*
40: qb_est21:*
41: qb_est22:*
42: qb_est23:*
43: qb_est24:*
44: qb_est25:*
45: qb_est26:*
46: qb_est27:*
47: qb_est28:*
48: qb_est29:*
49: qb_est30:*
50: qb_est31:*
51: qb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: qb_est33:*
60: qb_est34:*
61: qb_est35:*
62: qb_est36:*
63: qb_est37:*
64: qb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: qb_est39:*
70: qb_est40:*
71: qb_est41:*
72: qb_est42:*
73: qb_est43:*
74: qb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: qb_est45:*
80: qb_est46:*
81: qb_est47:*
82: em_est35:*
83: em_est36:*
84: em_est37:*
85: qb_est48:*
86: qb_est49:*
87: qb_est50:*
88: qb_est51:*
89: qb_est52:*
90: qb_est53:*
91: qb_est54:*
92: qb_est55:*
93: qb_qss1:*
94: qb_qss2:*
95: qb_qss3:*
96: qb_qss4:*
97: em_qss1:*
98: em_qss2:*
99: em_qss3:*
100: em_qss4:*
101: qb_qss5:*
102: qb_qss6:*
103: qb_qss7:*
104: qb_qss8:*
105: qb_qss9:*
106: em_qss5:*
107: em_qss6:*
108: em_qss7:*
109: em_qss8:*
110: em_qss9:*
111: em_qss10:*
112: em_qss11:*
113: qb_qss10:*
114: qb_qss11:*
115: em_qss12:*
116: qb_qss12:*

```
117:  qb_qss13:*
118:  qb_qss14:*
119:  qb_qss15:*
120:  qb_qss16:*
121:  qb_qss17:*
122:  qb_qss18:*
123:  qb_qss19:*
124:  em_qss13:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	53.8	5.1	1101	122	CNS0039G	AL063921 Drosophil
c 2	46.6	4.5	805	122	CNS009TB	AL053500 Drosophil
c 3	44.2	4.2	1101	122	CNS00000	AL065414 Drosophil
c 4	43.4	4.1	913	123	CNS01572	AL105680 Drosophil
c 5	43.2	4.1	1201	123	CNS016BI	AL106627 Drosophil
c 6	41.6	4.0	1101	122	CNS00LT2	AL078714 Drosophil
c 7	41.4	4.0	817	122	CNS0089K	AL051596 Drosophil
8	41.4	4.0	1201	122	CNS007E3	AL067050 Drosophil
9	40.4	3.9	286	81	C92687	C92687 C92687 Diet
10	40.4	3.9	511	80	C25641	C25641 C25641 Diet
c 11	40.4	3.9	1101	122	CNS00Y9S	AL097042 Drosophil
c 12	40.2	3.8	411	93	AQ008025	AQ008025 CIT HSP-2
13	40.2	3.8	878	123	CNS0187R	AL108993 Drosophil
14	40.2	3.8	1225	123	CNS0161D	AL106171 Drosophil
15	40	3.8	1101	122	CNS001LM	AL068473 Drosophil
c 16	39.8	3.8	582	48	AU053283	AU053283 AU053283
17	39.8	3.8	1101	122	CNS002AI	AL062343 Drosophil
c 18	39.6	3.8	1101	122	CNS003BD	AL064091 Drosophil
19	39.6	3.8	1201	123	CNS016FP	AL106687 Drosophil
c 20	39.2	3.7	836	122	CNS009GO	AL077933 Drosophil
21	39.2	3.7	1101	122	CNS00FMC	AL070972 Drosophil
c 22	39.2	3.7	1201	123	CNS0162F	AL106209 Drosophil
c 23	39	3.7	392	49	AI522064	AI522064 tl78c06.x
c 24	39	3.7	441	96	AQ303034	AQ303034 HS_3229_B
c 25	39	3.7	1069	122	CNS0107G	AL098614 Drosophil
c 26	39	3.7	1101	122	CNS0106X	AL098595 Drosophil
c 27	38.8	3.7	939	122	CNS00701	AL066258 Drosophil
c 28	38.6	3.7	426	25	AA364167	AA364167 EST74720
29	38.4	3.7	286	91	F03429	F03429 HSC1XA022 n
30	38.4	3.7	668	40	AI484706	AI484706 EST242967
c 31	38.4	3.7	1101	122	CNS000BJ	AL062049 Drosophil
c 32	38.2	3.6	332	42	AI678625	AI678625 tw84b04.x
c 33	38.2	3.6	403	42	AA918088	AA918088 oi81d06.s
c 34	38.2	3.6	426	32	AA862354	AA862354 wq94e10.s
c 35	38.2	3.6	518	69	AW235048	AW235048 xn18a06.x
c 36	38.2	3.6	660	119	AZ053289	AZ053289 RPCI-23 4
37	38	3.6	884	122	CNS007IY	AL067326 Drosophil
38	38	3.6	970	123	CNS0182E	AL108800 Drosophil
39	37.8	3.6	240	48	AU074260	AU074260 AU074260
c 40	37.8	3.6	332	27	AA483090	AA483090 ne71q06.s
41	37.8	3.6	445	88	R59083	R59083 yh03d11.s1
c 42	37.8	3.6	475	27	AA484732	AA484732 ne81104.s
43	37.8	3.6	491	88	R51299	R51299 yq72c07.s1
44	37.8	3.6	591	64	AW162119	AW162119 au73c12.x
c 45	37.8	3.6	932	123	CNS017PY	AL108352 Drosophil

ALIGNMENTS

RESULT 1
CNS0039G
LOCUS CNS0039G 1101 bp DNA GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL063921
VERSION AL063921.1 GI:4941778
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BACR Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI 98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR08K10"
/note="end : TET3"
BASE COUNT 301 a 64 c 131 g 202 t 503 others
ORIGIN

Query Match 5.1%; Score 53.8; DB 122; Length 1101;
Best Local Similarity 15.4%; Pred. No. 0.00021;
Matches 88; Conservative 248; Mismatches 237; Indels 0; Gaps 0;

```
Qy    321 tagtaatgcgatataatggccaaacaacgtactagaaatgactaataagtacccagtcaa 380
          ..|||    | | ||    |.|.||  ||    .:| : |  |:| :|   |  :||
Db    510 WWTTAAAWAAAAAAAAAAAAWAAAWAAAWATAAATWTWWTWWTTYTTWAAWATAAAMCMAA 569

Qy    381 gtcaactttgctgtagtattatatatttaacgaagcgtccattttactgccagggcaagtttat 440
          ...  .||  . .  ::|.  .... .  :   .  .... .|  ::    :: | :|:
Db    570 WYYHTYTTTYTYHYYTYWTYTMTWHYHTMYTHAWAHTTWYHWYHTYAMWHWMTWHWTMWAW 629

Qy    441 caatgggaccagtgttctcctcctctgaacaactcagttctttgcaaacgctagacagt 500
          ::|   | :  : | ::::::::::  |:|| :   :| :|: :::|  ::   :
Db    630 WHWTTYTAAYYYYYTCMYYYHYMHWHHAHAHAAAWWITHTWWTHAYHWAIYHYYYYMYCA 689

Qy    501 ctacctctctgccaccatttttacttttcaaaaatttactcctttgccgctactgaaaactt 560
          :  :||:||: |: ::::::| :: :|::  |:   :: :: ::: :|    :: ::::::
Db    690 MMCMCIHICHHCYYYYHHYIAHHIHTHHWYAHYYMWYYMWAYYWMYCTACTYHYHHHHHHY 749

Qy    561 ctacaattgaaagagccccacaatgaaagtccaagctacattcgccacccttatcgccctt 620
          ::|:: |: |:| |  : ::|:: ||| :  :|  ::::::|  :::: : :::: : | :
Db    750 HWAYHTTWYAWAHAMWMWHHAHYAAAAAWAAWATIHHYHHTTHYMHHIYMYHYYMYICCY 809

Qy    621 gcggcttacttccagcagccaatgcttgaaaagattgcatcatccaacgttataaaaac 680
          |  : :::::| :      :::   :|   :: :::   |::   : : :::::
Db    810 MCTYHCWHYYHTAYTCWTWTHHWMWTWTHWYHHTWWHHTTTHWAWWHTHTWCWWWWHATT 869

Qy    681 ggcgatgtcaacaacatatatactgccaataggaacgaaagataacctattgaggaatat 740
          : :  : :: ::: : ::: :  : :  :: :  : ::  |:::: :  :  ::
Db    870 WTWATHCWACMTMHWHHWMHMHHHHMACHAHHHTHMCMCHHHMHCTCHHHHTMYHMTCHW 929
```

```
Qy     741 aaaqtcttcqttaatqaqqcctqcqatcqcctaccaqqttatactttcccqacaqatqtcqttc 800
           ::: : ::: :::|| :  : | ::: |:::::|.. : .... .||. | .|: |:
Db     930 MWHMHWHWHHWWMATWMTTMTTMMMMCCMMHHHCHMYHMMHMYMYCCHYYCTCHTHATTH 989

Qy     801 ctttctqqccqattttacatcaqcttacqctqacqacqatqaqtcttqttqatcaataaqa 860
           ::: || : |:|| :: :||: :| | : :: : : | :: :: :|||: |
Db     990 YHYMCTCYHYCIWHIYWTAYWWAWIAHAMTTATWWWWMHWWAHWATWWWWWWWWATAWWA 1049

Qy     861 qtccaqqttqaaaaattcqccaccatqqtaata 893
           |||:  ::  :  ::::  ::: :    |:::
Db    1050 CTCHHWTWYHTHCTWYYHHTYHMWWAWWMAWHW 1082


RESULT   2
CNS009F8/c
LOCUS       CNS009F8       805 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR19C21 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL053500
VERSION     AL053500.1  GI:4934976
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 805)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : segret@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1..805
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib "RPCI-98"
                     /clone="BACR19C21"
                     /note="end : T7"
BASE COUNT       80 a      9 c    155 g    345 t    216 others
ORIGIN


  Query Match             4.5%;  Score 46.6;  DB 122;  Length 805;
  Best Local Similarity   36.1%;  Pred. No. 0.019;
  Matches   60;  Conservative   43;  Mismatches   63;  Indels    0;  Gaps    0;

Qy      44 cqccatatcttttttaaaatttaaccqattaaaaccqcaattcaattttqcqattttacaca 103
           |:|  :|| :|||: |: :|:|:|:  ::|| || |  :::   :  | :   || :
Db     716 CRCTAWTTTWTTTTTKAMTWTWTWAMTTYFAAAAAAMCTYYAYCCC*ACMCCYATAACTTW 657

Qy     104 aattaacaattcqtccaaaaqaqttatttaaqcqaagatttgtcatttttttaatccatt 163
           ||| || | ::  | |||| :    |  |  ::  |||:  ::||||: |:|  :::
Db     656 AATAAAAAYTWAMTATAAAAARAAATTCCTAHAARRATTTYSTTMYTTTTWSAWTAAMYT 597

Qy     164 tttaaaaaaatacatctgctttaacccacctttqccaaggqatqccc 209
           ||:||:|||  :   :|||||| ||: || :| :: : :  |:::
Db     596 TTWAAMAAAAATYRAAMK:::A:YCY:::MBTMMMSMCCNVAMMC


RESULT   3
CNS000D1/c
LOCUS       CNS000D1       [1:] bp    DNA             GSS       [illegible]
DEFINITION  Drosophila melanogaster genome survey seque[illegible]
            BACR01J16 of RPCI-98 library from Drosophi[illegible]
            fly), genomic survey sequence.
ACCESSION   AL065414
VERSION     AL065414.1  GI:4[illegible]
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; [illegible]
            Pteryqota; Neoptera; Endopteryqota; Diptera[illegible]
            Muscomorpha; Ephydroidea; Drosophilidae; D[illegible]
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre [illegible]
            BP 191 91006 EVRY cedex  FRANCE (E-mail : [illegible]
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was [illegible]
            collaboration with the Berkeley Drosophila[illegible]
            The BDGP is constructing a physical map of [illegible]
            melanogaster genome using these BACs. For [illegible]
            please see  http://www.fruitfly.org The BD[illegible]
            melanogaster BAC library was prepared by K[illegible]
            Aaron Mammoser in Pieter de Jong's laborat[illegible]
            Cancer Genetics at the Roswell Park Cancer[illegible]
            NY. The library is named RPCI-98 and was c[illegible]
            EcoRI digestion of Drosophila DNA provided[illegible]
            isogenic strain y2; cn bw sp, the same str[illegible]
            P1 and EST libraries. A more detailed desc[illegible]
            and how to order individual BAC clones, th[illegible]
            filters for hybridization from the BACPAC [illegible]
            found at http://bacpac.med.buffalo.edu/dro[illegible]
FEATURES             Location/Qualifiers
     source          1..1101
                     /organism="Drosophila melanogaste[illegible]
                     /db_xref="taxon:7227"
                     /clone_lib "RPCI-98"
                     /clone="BACR01J16"
                     /note="end : [illegible]
BASE COUNT       290 a    [illegible]  [illegible] g    211 t    [illegible]
ORIGIN


  Query Match             [illegible]%;  Score 44.2;  DB 122;
  Best Local Similarity   [illegible]%;  Pred. No. 0.095;
  Matches   65;  Conservative   [illegible];  Mismatches   [illegible];  [illegible];

Qy       2 atctqtaaattacatatatttatttqcqcatttqctaaacc[illegible]  [illegible]
           :::| |: :|| : [illegible]  : : :: :::::::  [illegible]
Db    1073 MHMIMIMTWTTTHTTHCHHMCMTTHCMTMTCYCWYMCHHHHHHMA[illegible]  [illegible]

Qy      62 ttttaacqattaaa[illegible]  [illegible]
           : :: | |:  : : :: : |:::: : | |:|:::::::  [illegible]
Db    1023 RCRHHYCTCttt[illegible]  [illegible]

Qy     122 aagaqqtatttaaqcqaaqatttqqcatttttttaatccatttt[illegible]  [illegible]
           : :  :  :| : :  : :  : | :::: |  :  | :  [illegible]
Db     953 YCMHCMYCWCMWATMWATWHA TMWHCHMIMHHHWCTAICIAW[illegible]  [illegible]

Qy     182 ctttaaccaccttt[illegible]  [illegible]
           ::  : |: ::: :: :  :  :  |: ::: :: :  [illegible]
Db     893 MWMMHTCTWTTWTHTC TCBMTHTTWTWTWITIWTMWHIAHYMW[illegible]  [illegible]

Qy     242 aaaqctqtcqatctatactcatttcqcqctctqtqaacaactaa[illegible]  [illegible]
           ::: |: :: ::: : : :::::  |:|:  :: :: :
Db     833 MHMICYAHMAHWTTTTTTHHHHHC CTCCHCYBCHCCCMHYM[illegible]  [illegible]
```

```
Qy      302 acacccctaaatcatgctgctagttatgcgatatgatgggccaaacaacgtactagaaatga 361
            |: ::: :   : :  :  ::   || :  :::::   ::||| ::: ||  | ::  :
Db      773 AYYYMHCWYCCYCYYTHHWHWTCTATMMMYWMWMHATTYMAAAAMMMATAAMAMMMMAHM 714

Qy      362 ctaataaqtaccccagtcaaqtcaacttgctgtagtattatatttaacgaaqcgtccattt 421
            ::::::: : :     :   :  ::|  : |:   |  |:  ||::  ||   | :: ||
Db      713 HHMWYMMTYTMTATMCYTCTCMTCMWTMAYHTMWMTTMCTHMMTAMMWAATAWTAHHATT 654

Qy      422 actgc 426
              :| |
Db      653 YMTMC 649


RESULT   4
CNS01572/c
LOCUS       CNS01572      913 bp    DNA             GSS       26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN13E21 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL105080
VERSION     AL105080.1  GI:5617094
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 913)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .913
                     /organism "Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN13E21"
                     /note="end : T7"
BASE COUNT      240 a    130 c    169 g    224 t    150 others
ORIGIN


  Query Match             4.1%;  Score 43.4;  DB 123;  Length 913;
  Best Local Similarity   37.3%;  Pred. No. 0.15;
  Matches   97;  Conservative   41;  Mismatches  122;  Indels    0;  Gaps    0;

Qy       49 attcttttaaaatttaacgattaaqaacgcaattcaatttgcgttctacacaaatta 108
            ::| |||:||:|:   ||   |:|||    |: ::| ||:| |   || | :|  || |
Db      881 HHTTTTTHTAWAMAACTABTTAYTAATKTAGVTMWTAAAHTGTAAATTATSSAATAAAAA 822

Qy      109 acaattcgtccaaaaqaqqtatttaaqcqaaqatttggcatttttttaatccatttttaa 168
             :  ||:      |: | :|:   :| ::||   :|:  : : ::  | |    |  |
Db      821 SVNTTTVAATKTTAVAAASTVCSSVACSVAAATAWTVATWSYSWWAAACTAAYCTMATYT 762

Qy      169 aaaaatacatctgctttaacccacctttgccaaggqtaccqggctagcatagccttggtt 228
            |:    |  | : :  |:|| |: |||    |  |  |   : |: ||:  |||    |
Db      761 AWTTYSATYTTWTSAATMACTAABTTTTATAKAWACTTTCAATSSASTAIWTTCTTCCST 702

Qy      229 accaaaaacqactaaaqctqtcqatctatactacattcqqcgctctqaacaactaaqcct 288
             :|||| |:  : ||:  | || |||| ||| || | : | : :  | ||| || | |
Db      701 TSCAAASASWASCAAWTATATCTATCTCAACTCCAATTSTCASAVAAATTAACCAAACTA 642
```

```
Qy      289 cag.gaggagatcacacccct 308
            :| |    | | ||| ||
Db      641 ATSCCAAYCCAACTCACWCT 622


RESULT   5
CNS016E1/c
LOCUS       CNS016E1      1201 bp    DNA             GSS       26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN15C13 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL106627
VERSION     AL106627.1  GI:5622848
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism "Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN15C13"
                     /note="end : SP6"
BASE COUNT      448 a    136 c    161 g    336 t    120 others
ORIGIN


  Query Match             4.1%;  Score 43.2;  DB 123;  Length 1201;
  Best Local Similarity   45.1%;  Pred. No. 0.18;
  Matches   74;  Conservative   19;  Mismatches   71;  Indels    0;  Gaps    0;

Qy        8 tctgtaaattacatatatttatttttgccgcattttgctaaccgcctattcttttaaaaat 67
            | | | |  || | | |: | :|||      ||||  ||| |  :|:|  | |:||:|:|
Db     1172 TTTTTTATATAAAAAAAWATTWTTTTTWTTTTTTTTTTAAWCTTYTWTATTATWTAWAWT 1113

Qy       68 tttaacgattaagaacgcaattcaatttgcgttctacacaaattaacaattcgtccaaa 122
            ||:|   || :| ||     || :|||:|||   || ||:  :||||   :|||  |   |:
Db     1112 TTWATTTATAWAWAAWATTATAHAAHTTTTTTTTTTIAMTAWAATTTTTWATTTTTTTAW 1053

Qy      123 aqaaqtatttaaqcqaaqatttqqcatttttttaatccatttttt 166
             :   ||||     :   |||    || :  ||:|   |||||
Db     1052 TIMNTNATTTTTWAAMTTTTTIATTTTTAWAATAWTTTTTTTTT 1009


RESULT   6
CNS00LT2/c
LOCUS       CNS00LT2      1101 bp    DNA             GSS       14-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC:
            BACR48P19 of RPC1-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL078714
VERSION     AL078714.1  GI:5102004
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
```

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (11-JUN-1999) Genoscope - Centre National de Sequencage , BP 191 91006 EVRY cedex FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see: http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://hgsc.bcm.tmc.edu/projects/drosophila/bac.html.
FEATURES Location/Qualifiers
source 1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR48P19"
/note="end : IET3"
BASE COUNT 469 a 6 c 69 g 151 t 406 others
ORIGIN

Query Match 4.0%; Score 41.6; DB 122; Length 1101;
Best Local Similarity 18.1%; Pred. No. 0.5;
Matches 88; Conservative 178; Mismatches 215; Indels 5; Gaps 1

Qy 258 actacattcaggcgctctgaacaactaaacaacaacgaaggaatcacacctaaatcatgc 317
Db 1101 MMMMMHHTTMMMMMMCMHTAAMMMAMMMMMMMMMMAMAMMMCMMCMCMCCMCMCHMMCAA 1042

Qy 318 tgctaqtaatgcgatataatggccaaacaacgctactagaaatgactaataagtaccccagt 377
Db 1041 MMMMHTHMMMMMMMMMAMMMMAMMMMMMMMMAITTTTHMMMAMAMMMHMMMMMMMMATTH 982

Qy 378 caagtcaacttgctgtaqtattatatttaacgaaqcgtccatttactgccagggcaaqtt 437
Db 981 AWHTTTTHTTTYMMAMCMTHTMMMMMMMMMAMMAMMMCCMCMCMMCCMMMCCCCMCMM 922

Qy 438 tatcaatgqgaccaqtqttctcgctcctctqqaacaactcaqttctttgcaaacgctaqac 497
Db 921 MMMCMMMITTTHHHMMCCMCMCMCCCMCMCMMCMMMACMMMMMTTTHMMCCCMMMMMMM 862

Qy 498 aqtctacctctctgccaccatttttacttttcaaaaatttactccttgccgctactgaaa 557
Db 861 MMMAAMMMAMMMMMCMHMAMMTTTTTTTTHWMAMMAYHTTMMMMTTMMMMMCMMMMAAA 802

Qy 558 cttctacaattgaaaqagccacaatgaaaqtccaaqctacattgccaccccttatgccc 617
Db 801 ATTMMMMMMAMHMTHHHCTMMMC-----MMCWCCMCMMMMMMMMCCMMCMMCMCMMMCCM 747

Qy 618 ctt[illegible] 677
Db 746 MYHMMMHTITWTTMMHMMMAMAIMMICHWMMTHMWMHTTTTMMMMMMAAAWWTTTWAHAAA 687

Qy 678 gacgqcgatqtcaacaacatatttattgccaatggaacgaqagatactattgaggaa 737
Db 686 WTTTTTTATTTAATAAAAAAAAAAAMAMTAAAAAAAAAAATWTAWAAAWAAAAATTWAAAAA 627

Qy 738 tataaa 743
Db 626 WAAAAA 621

RESULT 7
CNS0089K/c
LOCUS CNS0089K 817 bp DNA GSS
DEFINITION Drosophila melanogaster genome survey sequence BACR18E14 of RPCI-98 library from Drosophila (fruit fly), genomic survey sequence.
ACCESSION AL051596
VERSION AL051596.1 GI:4[illegible]
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; [illegible]
Pterygota; Neoptera; Endopterygota; Diptera;
Muscomorpha, Ephydroidea, Drosophilidae; D[illegible]
REFERENCE 1 (bases 1 to 817)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre [illegible]
BP 191 91006 EVRY cedex FRANCE (E-mail : [illegible]
Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was [illegible]
collaboration with the Berkeley Drosophila [illegible]
The BDGP is constructing a physical map of [illegible]
melanogaster genome using these BACs. For [illegible]
please see: http://www.fruitfly.org The BDG[illegible]
melanogaster BAC library was prepared by Ka[illegible]
Aaron Mammoser in Pieter de Jong's laborato[illegible]
Cancer Genetics at the Roswell Park Cancer [illegible]
NY. The library is named RPCI-98 and was c[illegible]
EcoRI digestion of Drosophila DNA provided [illegible]
isogenic strain y2; cn bw sp, the same str[illegible]
P1 and EST libraries. A more detailed descr[illegible]
and how to order individual BAC clones, the [illegible]
filters for hybridization from the BACPAC [illegible]
found at http://hgsc.bcm.tmc.edu/projects/dros[illegible]
FEATURES Location/Qualifiers
source 1..817
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR18E14"
/note="end : IET3"
BASE COUNT 241 a [illegible] 154 g 227 t
ORIGIN

Query Match 4.0%; Score 41.4; DB 122;
Best Local Similarity 42.9%; Pred. No. 0.52;
Matches 96; Conservative 19; Mismatches 111; [illegible]

Qy 20 tttatttgcgcgcatttgctaaccgctctattcttttaaaattt [illegible] 79
Db 633 TTTTTTTTTWTTTTTAAATAAATTTCTTYTTTTTTTTTTTW [illegible] 573

Qy 80 caattcaatttgtgttctacgaaattaacaattcgtccaaaa [illegible] 129
Db 573 TAAGYTTATCTAGAATTCTAAGYKTTKTTTCAAAGCCTTTCAA [illegible] 523

Qy 140 gatttggcatttttttaatccatttttaaaaaaatatcatctgct [illegible] 189
Db 523 AATTKTGWTWTWTTTTTAAAAATKTKTTTTTTTANAACCGAAGCA [illegible] 473

Qy 200 aagggtaccgggctaccataqccttqgttaccaaaaacgctaa [illegible] 249
Db 453 AWTAKTACACGAAA[illegible]TAANTGAGGTAAATTTTTNTA [illegible] 403

RESULT 8
CNS007E3
LOCUS CNS007E3 1291 bp DNA G[illegible]

```
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR15H14 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL067050
VERSION     AL067050.1  GI:4945614
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope  Centre National de Sequencage .
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org/The_BDGP_Drosophila
            melanogaster_BAC_library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            PI and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/Drosophila_bac.htm
FEATURES             Location/Qualifiers
     source          1..1201
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR15H14"
                     /note="end : TET3"
BASE COUNT      266 a     98 c     77 g    586 t    174 others
ORIGIN


  Query Match             4.0%;  Score 41.4;  DB 122;  Length 1201;
  Best Local Similarity   42.9%;  Pred. No. 0.58;
  Matches   75;  Conservative   23;  Mismatches   77;  Indels    0,  Gaps    0;

Qy      11 ttacatatatttatttttqccqcatttqctaaccgcctatttcttttaaaatttaacqa 70
           ::| || | ||| |:: : |   |||| || : | : :|||   ||||
Db     793 WYATATTTTTTTTTBWCDACTTTTTTTTTTAWTHACACGTYTYTTTTTTTATTTTTATT 852

Qy      71 ttaaaaacqcaattcaattttqcattctacaaaattaacaattcqtccaaaaqaqqtat 130
           |   :||    |: : |||::   || || | |  | ||:     | | |     |:
Db     853 TAYTSAAACAGAKAYTATTKWTTTTTTTATATATTTASACWCGATWTTTACAHAYTTTWA 912

Qy     131 ttaaqcgaagatttggcatttttttaatccatttttaaaaaaatacatctqctttt 185
           ||| :|:|:|:    :||||||  |   |||||  : ||||   :|| ||||
Db     913 AVAAGARAAKAKTYTTTTWTTTTTTTTTTTTTTTTTYTWCAAATYACKCTTCTTT 967


RESULT  9
C92687
LOCUS       C92687          286 bp    mRNA            EST       12-JUL-1999
DEFINITION  C92687 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
            discoideum cDNA clone SSF838, mRNA sequence.
ACCESSION   C92687
VERSION     C92687.1  GI:3074563
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Dictyostelilda; Dictyostelium.
REFERENCE   1  (bases 1 to 286)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     The Dictyostelium developmental cDNA project: generation and
            analysis of expressed sequence tags from the first-finger stage of
            development
  JOURNAL   DNA Res. 5 (6), 335-340 (1998)
  MEDLINE   99156227
COMMENT     On Jan 6, 2000 this sequence version replaced gi:6675731.
            Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
            Email: d402hu@sakura.cc.tsukuba.ac.jp
            PROJECT - 'Dictyostelium discoideum cDNA project in Japan'
            POLYA-No.
FEATURES             Location/Qualifiers
     source          1..286
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SSF838"
                     /clone_lib="Dictyostelium discoideum SS (H.Urushihara)"
                     /dev_stage="slug"
BASE COUNT      128 a     20 c      5 g    133 t
ORIGIN


  Query Match             3.9%;  Score 40.4;  DB 81;  Length 286;
  Best Local Similarity   53.1%;  Pred. No. 0.76;
  Matches   86,  Conservative    0,  Mismatches   76,  Indels    0;  Gaps    0;

Qy       9 aaattacatatatttattttttcccqcatttgctaaccgcctatttcttttaaaatttaaa 67
           || ||    | | |||||||      | ||  | |    | |||  | |||| |||||||
Db      87 AATTTTTTGAAAATTATTTTTATTATTATTTATTATTAGCAATTAATATTAATATTTTAA 146

Qy      68 cgattaagaacgcaattcaatttttgcgttctacacaaattaacaattcgtccaaaaqaqg 127
           |||| |    | || ||||||   || || | |||  || ||     | |||| |
Db     147 TTTTTAAAATTAAATTTTAATTTTAATTTTTAAAAAAAAAAAAAAAAAAACTAAAAAAAA 206

Qy     128 tatttaagcgaagatttggcatttttttaatccatttttaaa 169
           |   |||  | |      ||| |||| |||   |   |||
Db     207 AAAAAAAGAAACAAACAATTATTATTTTTATCACATCAAAAA 248


RESULT  10
C25641
LOCUS       C25641          511 bp    mRNA            EST       28-APR-1999
DEFINITION  C25641 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
            discoideum cDNA clone SSA838, mRNA sequence.
ACCESSION   C25641
VERSION     C25641.1  GI:2281857
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Dictyostelilda; Dictyostelium.
REFERENCE   1  (bases 1 to 511)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     The Dictyostelium developmental cDNA project: generation and
            analysis of expressed sequence tags from the first-finger stage of
            development
  JOURNAL   DNA Res. 5 (6), 335-340 (1998)
  MEDLINE   99156227
COMMENT     On Jul 7, 1999 this sequence version replaced gi:5866374.
            Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
            Email: d402hu@sakura.cc.tsukuba.ac.jp
            PROJECT - 'Dictyostelium discoideum cDNA project in Japan'
            POLYA-No.
FEATURES             Location/Qualifiers
```

```
     source          1. .511
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SSA838"
                     /clone_lib "Dictyostelium discoideum SS (H.Urushihara)"
                     /dev_stage="sing"
BASE COUNT      193 a     70 c     54 g    182 t     12 others
ORIGIN


  Query Match             3.9%;  Score 40.4;  DB 80;  Length 511;
  Best Local Similarity   53.1%;  Pred. No. 0.88;
  Matches   86;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy       8 aaattacatatatttatttqccqcattttqctaaccqcctattctttttaaaattttaa 67
           || ||    | | |||||||        | ||  | |    | |||  | |||| |||||||
Db     316 AATTTTTTGAAAATTATTTTATTATTATTTATTATTAGCAATTAATATTAATATTTTAA 375

Qy      68 cgattaaqaaccqcaattcaatttqcgttctacacaaatttaacaattcqtccaaaaqaqq 127
           |||| |    | || ||||||   || || | |||  || ||     | |||| |
Db     376 TTTTTAAAATTAAATTTTAATTTTAATTTTTAAAAAAAAAAAAAAAAAAACTAAAAAAAA 435

Qy     128 tatttaaqccaaaqatttqqcatttttttaatccatttttaaa 169
           |   |||  |  |      ||| |||| |||   |   |||
Db     436 AAAAAAAGAAACAAACAATTATTATTTTTATCACATCAAAAA 477


RESULT  11
CNS00YZS/c
LOCUS       CNS00YZS     1101 bp    DNA             GSS       26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence 3'6 end of BAC
            BACN01M04 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL097042
VERSION     AL097042.1  GI:5608653
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pteryqota; Neoptera; Endopteryqota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUL-1999) Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payen. It has been constructed in the vector
            pBeloBAC11
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib "DrosBAC"
                     /clone="BACN01M04"
                     /note="end : SP6"
BASE COUNT      292 a    179 c    182 g    341 t    107 others
ORIGIN


  Query Match             3.9%;  Score 40.4;  DB 122;  Length 1101;
  Best Local Similarity   45.8%;  Pred. No. 1.1;
  Matches   88;  Conservative   20;  Mismatches   80;  Indels    4;  Gaps    1;

Qy       2 atctqtaaattacatatatttatttqccqcattttqctaaccqcctattctttttaaaa 61
           || | |||  :| | |||:.  | |       | ||  :|
Db     991 ATTTTAAtNWAAAAATWtAtttttAWNWNtNttttWAAAAAt                 944

Qy      62 tttaaccqattaaqaaccqattcaatttqcqttctacacaaatt                111
           ||: ||  | ||| :|   | |     | :|   :| | |  : |
Db     934 TTWAAATAAATAAAAAAAAAtt:   t:WttWAWtWttWAWWttt              886

Qy     122 aaqaaqtatttaaaccqaaaatttaacaattttttttaatccatttttt           171
           || |  | |||||   :: :        ||||||||   |:  :
Db     878 AAWATTTTTTTAATtAAWtAAAAAAtt:tttttttAAAttAKAAtt            829

Qy     182 ctttaaccccacc 195
           |   ::::::::
Db     818 CMMMMMMMMMMM 807


RESULT  12
AQ008025/c
LOCUS       AQ008025      411 bp    DNA             GSS       26-JUL-1999
DEFINITION  CIT-HSP-2296J21.18 TJ1 HSP Homo sapiens ge
            genomic survey sequence.
ACCESSION   AQ008025

VERSION     AQ008025.1  GI:5008025
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Ve
            Mammalia; Eutheria; Primates; Catarrhini;
REFERENCE   1  (bases 1 to 411)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Zhao,S., Fleid,
            Golden,K., Berry,K., Fraser,C.M., Suhre
            Simon,M. and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database
            Building (1998)
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 208
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetic
            end search page:
            http://www.tigr.org/8_forms/bac_req.html
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .411
                     /organism="Homo sapiens"
                     /db_xref="GDB:7153033"
                     /db_xref="taxon:9606"
                     /clone="2296J21"
                     /clone_lib "CIT-HSP"
                     /sex="Male"
                     /cell_type="Sperm"
                     /note="Vector: pBeloBAC11; Site_
                     HindIII"
BASE COUNT      192 a     64 c     64 g     91 t
ORIGIN


  Query Match             3.8%;  Score 40.2;  DB 93;
  Best Local Similarity   51.4%;  Pred. No. 0.94;
  Matches   93;  Conservative    0;  Mismatches   88;                Gaps

Qy      44 cqcctattcttttaaaattttaaccqattaaqaaccqcaattcaa             93
           | ||| || | ||  ||||| | |||  || | | |  : |
Db     245 CTCCTTTTATGTTtCTAAtTTtAATTATTGGAtCtCtCTtCtCt             196

Qy     104 aattaacaattcqtccaaaaaaqaqtatttaaaccqaaaqatttqq             153
           |    || || :                 || |  | | | ::
```

```
Db     185 ACGGTTCATCTCTTTTCTTTATCTTTTTAGAAAAACAACTTTTCATTTTATTCATCTTTT 126

Qy     164 tttaaaaaaatacatctgctttaacccacctttgccaaggtaccctgctagcat 223
           ||    | || |||     ||||       |||  ||| |     |    ||  | ||| |
Db     125 CTTTGGAGAAGACAAGATGTTTATTGGTGCTTCTCCATGCACCTCTATTTATGAAAGCAT 66

Qy     224 t 224
           |
Db      65 T 65


RESULT  13
CNS0187R
LOCUS       CNS0187R      878 bp    DNA             GSS       26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN04E04 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL108993
VERSION     AL108993.1  GI:5629297
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 878)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .878
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib "DrosBAC"
                     /clone="BACN04E04"
                     /note="end : SP6"
BASE COUNT      279 a    132 c    120 g    182 t    165 others
ORIGIN


  Query Match              3.8%;  Score 40.2;  DB 123;  Length 878;
  Best Local Similarity   23.1%;  Pred. No. 1.1;
  Matches   63;  Conservative   86;  Mismatches  124;  Indels    0;  Gaps    0;

Qy      99 acacaaattaacaattcgtcccaaaagaggtatttaagcgaagatttgcgcatttttttaat 158
           | | :::  || ||    :::::|| |   :  :| : :| |     :       :|:
Db      64 AAAAWMWMCAAAAAAAAAAMMWMAAAAAWAWAAAMAWMWMAAAAAAAWWWAMARGAWAWG 123

Qy     159 ccatttttaaaaaaatacatctgctttaacccacctttgccaaggtaccctgctagcat 218
           | :   :::::|::  :::: :     ||  :: |||||:  |  |:      |: |:
Db     124 CWWGGAWWMMMAWRGMMMMWAWWAAAAAAWWMWWCTTTGCMCTGAWTRTWGTTRTWTWAW 183

Qy     219 agcctgqttaccaaaaacgqctaaaqctqtcgatctatactacattcgqcgctctgaac 278
           : : ||:  :: :||::::   :::|: : : :  : | |  : :  : :|:|: : : :
Db     184 WWMRTTRAAWWWMAAWWWMAAWWWARGWWWWRGMMAAAARAWWWAMMWRCRCWTWCRCM 243

Qy     279 aactaaqcaacaqcqaqqaqatcacaccctaaatcatgctgctagtaatgcgatataatq 338
           ||: ||  || |   :  | | || : :  ||:::|:    |  :  ||:  ||:::::
Db     244 AAMMAAMAAAAAMAAMMAAMARCAWMWMWMAAMWCRAAWMGGAWARAAWTRGAWWWMRAM 303

Qy     339 qccaaacaacgctactaqaaatgactaataaqta 371
              |:: |:: :: :: ||| : :  :|| :|
Db     304 MAAAMKAARMAWRAWMMAAAWAMAATWTACAWA 336


RESULT  14
CNS0161D
LOCUS       CNS0161D      1225 bp    DNA             GSS       26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN15C18 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL106171
VERSION     AL106171.1  GI:5620504
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1225)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1225
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN15C18"
                     /note="end : SP6"
BASE COUNT      266 a    128 c     38 g    368 t    425 others
ORIGIN


  Query Match              3.8%;  Score 40.2;  DB 123;  Length 1225;
  Best Local Similarity   38.1%;  Pred. No. 1.2;
  Matches   69;  Conservative   32;  Mismatches   80;  Indels    0;  Gaps    0;

Qy       7 taaattacatatatttattttgccqcattttgctaaccqcctattcttttttaaaatttta 66
           |||:|||  | | ||||:::|     : |    ||      | || |||:||||:   |
Db     968 TAAWTTATTTTTTTTTAWWWTTTATWWATAATTTATWTTTWTTTTTTTTWTAAAAWAAAA 1027

Qy      67 acgattaaqaaaqcaattcaattttgcqttctacacaaattaacaattcqtccaaaaqaq 126
           |  | ||        : : : :|:    || || | : |:|   |  |  :  |:|| :
Db    1028 AAAAATATATTTTWTTAWWWTWTWATTTTTATATATWTAYTTTTATATTMWAWAWWATWW 1087

Qy     127 gtatttaaqcgaagatttgcgcatttttttaatcaattttaaaaaaatacatctgcttta 186
           |:|:|:     : :|::   || |||||: | ||:| :|| :| ||| |  :  |||:
Db    1088 TTWTWTWTTWTTWWWTWWWTAATATITTTWTTTTATWTAWAATWATATAAAAAWTWTTTW 1147

Qy     187 a 187
           :
Db    1148 W 1148


RESULT  15
CNS00LLM
LOCUS       CNS00LLM     1101 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7T3 end of BAC:
            BACR30I18 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL068473
VERSION     AL068473.1  GI:4958603
KEYWORDS    GSS.
```

```
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera,
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1..1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR30I18"
                     /note="end : TET3"
BASE COUNT       387 a    157 c     27 g    186 t    344 others
ORIGIN
```

```
  Query Match              3.8%;  Score 40;  DB 122;  Length 1101;
  Best Local Similarity   25.7%;  Pred. No. 1.4;
  Matches  118;  Conservative  119;  Mismatches  219;  Indels    4;  Gaps    1;

Qy     158 tccatttttaaaaaaatacatctgctttaacccacctttgccaagggtaccccgctagca 217
           ||||   |   |   :  |: || : |   :|| |:| : :::    || |  ::|  :
Db     536 TCCAAACIMCYAYCWCAAHCTCHCYCTAYTHCCYCHTAYAHMMYMTWCACACTTMYAYTM 595

Qy     218 tagccttgattacccaaaaaaggctaaaictatcgatctatactacattccgagctttgaa 277
           |  |:|   ||: : :::   ::||: :|    :  ::||| :   :   | ||:   |
Db     596 NACICWIMIMIAMNMIWMAAWAYHAAWAHIAATAWCWYHIACAHACAHTYICICIMCWTA 655

Qy     278 caactaagcaacaacgcgaggaagatcacaccctaaaatcatgctgctaqtaatgccgatataat 337
           :| |  | |: :|   |  | :|:   : ::| ||       :  |:|         |:
Db     656 YATCATAMCMTMAHATAATATMTMTATMTMHACATTYNAAATAYTHTWACTTATCTMNAH 715

Qy     338 qgccaaacaacgtactatgaaatgactaataagtaccccagtcaagtcaacttgctgtagta 397
           :|:|||    | :| ||:| ::|:: :: :|: :| |:   |: | :|     :  :
Db     716 IMWMAMACATAACAAHAYAAHIMWYIMWMWMMYAHTHAWIMYNCIMIAWWIMIAACWAAW 775

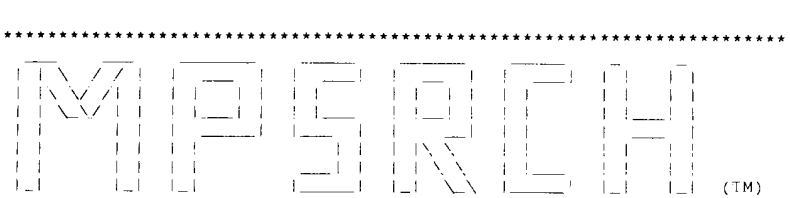
Qy     398 ttatatttaacgaaagctccatttactgccaggacaagtttatcaatggcaccagtgttc 457
           | :|||    || : : ||| |:|   :        |: :|   :     :| | ||:
Db     776 AAAAHITTTAYAATMTHICATMTWCNAIYTAWMNTTHTWNHTACTYTYICTYATTWTTH 835

Qy     458 tccctc----ctctgaacaactcagttctttgcaaacgctagacagctctcctctctgcc 513
           |:|: :    ||:  ::|:  :  : :|    |:: :|: :|: :: :  | :  |:
Db     836 TYCYNMNMTICTYCCTMYAMTMIMIMHTWTACWTAMMYMIMHMCMMHMAYMTMCCMCACM 895

Qy     514 accatttttacttttcaaaaatttactccttgccgctactgaaacttctacaattgaaag 573
           |||| | |:: | : :: :|    ::  | |  ::       |: |:|  :||| ::
Db     896 CTCAIMTATAYHCICYIWHTMICNCIWYIMIYCAWYHNWCMICAYMICICTMATIAHWCT 955

Qy     574 agcccacaatgaaagtccaagctacattcgccaccctat 613
           :  |:|: :| ::: :|: : ||:: | : ::  |  |::
Db     956 HIACMAYNMIMMMCYCMCHICIMYCICYCYMYICIMIMW 995
```

```
Search completed: August 27, 2000, 22:33:06
Job time: 18160 sec
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993 1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 25 17:53:48 2000; MasPar time 10.75 Seconds
574.123 Million cell updates/sec

Tabular output not generated.

Title: >US-09-257-585-4
Description: (1-89) from US09257585.pep
Perfect Score: 634
Sequence: 1 MKVQAIFAFLIALAAYFPAA.....DRSVLSGDFISAYADDDESC 89

Scoring table: PAM 150
Gap 11

Searched: 225878 seqs, 69344122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl12
1:sp_archea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 38.662; Variance 66.101; scale 0.585

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	93	14.7	723	5	O45470	F36G9.12 PROTEIN.	8.43e-02
2	92	14.5	637	14	O72407	PF D11.	1.20e-01
3	91	14.4	1275	4	Q15057	KIAA0349 (FRAGMENT).	1.71e-01
4	91	14.4	1353	5	Q18419	C34B4.2 PROTEIN.	1.71e-01
5	91	14.4	1661	5	Q06166	MATURE PARASITE-INFECT	1.71e-01
6	90	14.2	1201	6	O18978	NA-K-CL COTRANSPORTER.	2.44e-01
7	90	14.2	1510	5	Q25920	MATURE-PARASITE-INFECT	2.44e-01
8	89	14.0	353	2	Q9ZNH0	3-HYDROXY-3-METHYLGLUT	3.46e-01
9	88	13.9	875	5	Q19045	E03A3.6 PROTEIN.	4.91e-01
10	87	13.7	219	14	O72729	D6L PROTEIN.	6.93e-01
11	87	13.7	301	1	O57964	301AA LONG HYPOTHETICA	6.93e-01
12	87	13.7	326	3	Q9Y8C9	MATING TYPE PROTEIN MA	6.93e-01
13	87	13.7	391	1	O58489	391AA LONG HYPOTHETICA	6.93e-01
14	87	13.7	592	10	Q9ZPS7	PUTATIVE MULTISPANNING	6.93e-01
15	87	13.7	720	2	Q51162	FE-REGULATED PROTEIN B	6.93e-01
16	87	13.7	922	2	P94593	YWQA PROTEIN (MEMBER O	6.93e-01
17	87	13.7	4017	14	Q88508	(PURDUE-115).	6.93e-01
18	86	13.6	192	1	O59618	192AA LONG HYPOTHETICA	9.76e-01
19	86	13.6	728	5	Q9XXG4	Y37A1B.14 PROTEIN.	9.76e-01
20	85	13.4	344	5	Q9XTB9	H22D14.1 PROTEIN.	1.37e+00
21	85	13.4	394	5	Q19438	F14A5.1 PROTE	1.37e+00
22	85	13.4	402	10	Q47228	GLUCOSE-6-PH	1.37e+00
23	85	13.4	868	4	O00602	TOLL/INTERLE	1.37e+00
24	84	13.2	448	2	Q99495	PUTATIVE MEM	1.92e+00
25	84	13.2	789	5	Q01825	CODED FOR P	1.92e+00
26	83	13.1	167	5	Q2409	GLOBIN XII	2.64e+00
27	83	13.1	315	5	Q07494	PPG10-W PR	2.64e+00
28	83	13.1	376	1	Q08680	ORF YOR194W	2.64e+00
29	82	12.9	296	1	Q11076	ZAG1.	3.61e+00
30	82	12.9	342	5	Q18709	SIMILAR T	3.61e+00
31	82	12.9	508	5	Q19423	SIMILAR T	3.61e+00
32	82	12.9	617	1	O54654	617AA LONG	3.61e+00
33	81	12.8	134	10	Q80767	TI4(06.5).	4.94e+00
34	81	12.8	184	14	Q41204	GLYCOPROT	4.94e+00
35	81	12.8	384	14	Q41203	GLYCOPROT	4.94e+00
36	81	12.8	438	2	Q95508	HYPOTHETI	4.94e+00
37	81	12.8	2400	14	Q86085	S PROTEIN	4.94e+00
38	81	12.8	386	2	Q9ZIA1	HYPOTHET	4.94e+00
39	81	12.8	403	2	Q07563	HYPOTHET	4.94e+00
40	81	12.8	1449	14	Q9YRA7	VIRUS PO	4.94e+00
41	81	12.8	1449	14	Q88510	VIRUS PO	4.94e+00
42	81	12.8	1449	14	Q84853	RNA. O	4.94e+00
43	80	12.6	179	14	Q9W922	IRYMIT	6.78e+00
44	80	12.6	422	5	Q25673	ERYTHR	6.78e+00
45	80	12.6	718	14	Q88418	SPVI	6.78e+00

ALIGNMENTS

RESULT 1
ID O45470 PRELIMINARY; PRT; 723 AA
AC O45470;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01 JUN 1998 (TrEMBLrel. 06, Last sequence
DT 01-JAN-1999 (TrEMBLrel. 09, Last annotatio
DE F36G9.12 PROTEIN.
GN F36G9.12.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea
OC Rhabditina; Rhabditoidea; Rhabditidae; P
RN [1]
RP SEQUENCE FROM N.A.
RA WALL M.;
RL Submitted (NOV 1997) to the EMBL/GenBank
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE, 94158...
RA WILSON R., AINSCOUGH R., ANDERSON K., B
RA BONFIELD J., BURTON J., CONNELL M., CO
RA CRAXTON M., DEAR S., DU Z., DURBIN R.,
RA GARDNER A., GREEN P., HAWKINS T., HIL
RA JONES M., KERSHAW J., KIRSTEN J., LAI
RA LIGHTNING J., LLOYD C., MCMURRAY A., M
RA PARSONS J., PERRY C., RIFKEN L., ROGE
RA SMALDON N., SMITH A., SONNHAMMER E.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M.
RA WATSON A., WEDGWOOD R., WILKINSON-S
RT "2.2 Mb of contiguous nucleotide seq
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL; Z81543; CAB04331.1; -.
SQ SEQUENCE 723 AA; 82684 MW; E9C

Query Match 14.7%; Score
Best Local Similarity 32.4%; Pred.
Matches 11; Conservative 11; M

Db 1 MNPLISTLFLIQLGNYFANAKEWKEILL
Qy 1 MKVQAIFAFLIALAAYFPAANAWKEILL

RESULT

```
ID   O72907      PRELIMINARY;      PRT:  637 AA.
AC   O72907;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   FP-D11.
GN   FP-D11.
OS   Fowlpox virus.
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Avipoxvirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HP-440;
RA   SKINNER M.A.;
RL   Submitted (MAY 1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ005163; CAA06401.1; -.
DR   PFAM; PF00271; helicase_C; 1.
DR   PFAM; PF00176; SNF2_N; 1.
SQ   SEQUENCE   637 AA;  72774 MW;  DC5B6DEF CRC32;

  Query Match           14.5%;  Score 92;  DB 14;  Length 637;
  Best Local Similarity  31.0%;  Pred. No. 1.20e-01;
  Matches   13;  Conservative   11;  Mismatches  15;  Indels   3;  Gaps   3;

Db    108 KDCIIMNYDDQNFHNKFFTNIKS-INVKS-RIFIIIDECHNF 147
          ||||| | | : :| : :| :  | :  ::|:  : || :
Qy     24 KDCIIQPYKPCPVNNIYTANENEEITEEYKVFVN EACHPY 64


RESULT    3
ID   O15057      PRELIMINARY;      PRT: 1275 AA.
AC   O15057;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   KIAA0349 (FRAGMENT).
GN   KIAA0349.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 97349984.
RA   NAGASE T., ISHIKAWA K., NAKAJIMA D., OHIRA M., SEKI N., MIYAJIMA N.,
RA   TANAKA A., KOTANI H., NOMURA N., OHARA O.;
RT   "Prediction of the coding sequences of unidentified human genes. VII.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 4:141-150(1997).
DR   EMBL; AB002347; BAA20806.1; -.
FT   NON_TER       1       1
SQ   SEQUENCE   1275 AA;  145331 MW;  31D4F09F CRC32;

  Query Match           14.4%;  Score 91;  DB 4;  Length 1275;
  Best Local Similarity  26.7%;  Pred. No. 1.71e-01;
  Matches   12;  Conservative   15;  Mismatches  15;  Indels   3;  Gaps   3;

Db   1092 ESWCRNSEVKRYLEGERDAIKYPRESNKLINLPEDYSSLINQASN 1136
          ::| ::  : || :|: : | |  : |  | : |:|  ::|:| :
Qy     21 NAW-KDCIIQPYKPCPVNNI-YTANPNPEITI-EEYKVFVNEACH 62


RESULT    4
ID   Q18419      PRELIMINARY;      PRT: 1353 AA.
AC   Q18419;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   C34B4.2 PROTEIN.
GN   C34B4.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   SMYE R.;
RL   Submitted (AUG 1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SADDY D., SCOTHERN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z78059; CAB01486.1; -.
DR   PFAM; PF01064; Activin_recp; 1.
SQ   SEQUENCE   1353 AA;  153486 MW;  C50132E2 CRC32;

  Query Match           14.4%;  Score 91;  DB 5;  Length 1353;
  Best Local Similarity  28.9%;  Pred. No. 1.71e-01;
  Matches   13;  Conservative   12;  Mismatches  17;  Indels   3;  Gaps   3;

Db    973 YKQNYMTNETYVDEGKPFSVE-CQGHFCVIVPNKYIIDGEYVDIY 1016
          | |   :|  ::| | |  | |:  : ||:|:: :: |::   |
Qy     40 YTAN-RNEEITIEEYKVFVNEACH-PYPVILPDRSVLSGDFTSAY 82


RESULT    5
ID   Q06166      PRELIMINARY;      PRT: 1661 AA.
AC   Q06166;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   MATURE PARASITE-INFECTED ERYTHROCYTE SURFACE ANTIGEN
DE   (ANTIGENIC PROTEIN PFEMP2).
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PALO ALTO;
RX   MEDLINE; 92158014.
RA   COPPEL R.L.;
RT   "Repeat structures in a Plasmodium falciparum protein (MESA) that
RT   binds human erythrocyte protein 4.1.";
RL   Mol. Biochem. Parasitol. 50:335-347(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PALO ALTO;
RA   KUN J.F.J., WALLER K.L., COPPEL R.L.;
RL   Submitted (APR 1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 797-850 FROM N.A.
RX   MEDLINE; 93122844.
RA   SAUL A., YEGANEH F., HOWARD R.J.;
RT   "Conservation of repeating structures in the PfEMP2/MESA protein of
RT   Plasmodium falciparum.";
RL   Immunol. Cell Biol. 70:353-355(1992).
DR   EMBL; AF056936; AAC13303.1; -.
DR   EMBL; S52458; AAB24869.1; -.
DR   HSSP; Q57997; 1MJH.
KW   Antigen.
SQ   SEQUENCE   1661 AA;  195479 MW;  6FFCDF5E CRC32;

  Query Match           14.4%;  Score 91;  DB 5;  Length 1661;
  Best Local Similarity  24.6%;  Pred. No. 1.71e-01;
  Matches   15;  Conservative   20;  Mismatches  22;  Indels   4;  Gaps   4;
```

```
Db    1508 RRNCKKEKKYNNGELEEYYKEKQNEEYFDEEY-I-TQSKVHNTLNIFPNMA-LNEDFRRE 1564
           :::|  :: |:::|:::: |  ::|||   ||| : :    |     :|: : |: ||
Qy      23 WKDCTIQR-YKDGDVNNIYTANRNEFITTEEYKVFVNEACHPYPVILPDPSVLSGDFTSA 81

Db    1565 F 1565
           :
Qy      82 Y 82


RESULT    6
ID    O18978       PRELIMINARY;       PRT;  1201 AA.
AC    O18978;
DT    01-JAN-1998 (TrEMBLrel. 05, Created)
DT    01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT    01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE    NA-K-CL COTRANSPORTER.
OS    Bos taurus (Bovine).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC    Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea; Bovidae;
OC    Bovinae; Bos.
RN    [1]
RP    SEQUENCE FROM N.A.
RC    TISSUE=AORTA;
RA    YERBY T.R., VIBAT C.R.T., SUN D., PAYNE J.A., O'DONNELL M.E.;
RL    Am. J. Physiol. 274:Q-Q(1997).
DR    EMBL; U70138; AAC48754.1; -.
DR    PFAM; PF00324; aa_permeases; 1
SQ    SEQUENCE   1201 AA;  130393 MW;  CA143E11 CRC32;

  Query Match           14.2%;  Score 90;  DB 6;  Length 1201;
  Best Local Similarity  25.5%;  Pred. No. 2.44e-01;
  Matches   13;  Conservative   13;  Mismatches  23;  Indels   2;  Gaps   2;

Db    1024 LTILIPYLLITKKKWKDCKIPVFICCKINPIDHDRRAMPTLLIKFRIDFSD 1074
           |  | :|: :::  ||||  |: :  | :| |    |     :  ::: | :
Qy      10 LIALAAYFPAANA-WKDCTIQRYKDGDVNNIYTANRNEFITTEEYKV-FVN 58


RESULT    7
ID    Q25920       PRELIMINARY;       PRT;  1510 AA.
AC    Q25920;
DT    01-NOV-1996 (TrEMBLrel. 01, Created)
DT    01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT    01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE    MATURE-PARASITE-INFECTED ERYTHROCYTE SURFACE ANTIGEN.
GN    MESA.
OS    Plasmodium falciparum.
OC    Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN    [1]
RP    SEQUENCE FROM N.A.
RC    STRAIN=D10;
RX    MEDLINE; 92158014.
RA    COPPEL R.;
RT    "Repeat structures in a Plasmodium falciparum protein (MESA) that
RT    binds human erythrocyte protein 4.1.";
RL    Mol. Biochem. Parasitol. 50:335-347(1992).
DR    EMBL; M69183; AAA29651.1; -.
DR    HSSP; Q57997; 1MJH.
SQ    SEQUENCE   1510 AA;  177185 MW;  492B0D0A CRC32;

  Query Match           14.2%;  Score 90;  DB 5;  Length 1510;
  Best Local Similarity  24.6%;  Pred. No. 2.44e-01;
  Matches   15;  Conservative   21;  Mismatches  21;  Indels   4;  Gaps   4;

Db    1357 RRNCKKEKKYNNGELEEYYKEKQNEEYFDEEY-I-TQSKVHNTLNIFPNMA-LNEDFRRE 1413
           :::|  :: |:::|:::: |  ::|||   ||| :: :    |     :|: : |: ||
Qy      23 WKDCTIQR-YKDGDVNNIYTANRNEFITTEEYKVFVNEACHPYPVILPDPSVLSGDFTSA 81

Db    1414 F 1414
           :
Qy      82 Y 82


RESULT    8
ID    Q92NH0       PRELIMINARY;       PRT;  353 AA.
AC    Q92NH0;
DT    01-MAY-1999 (TrEMBLrel. 10, Created)
DT    01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT    01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE    3-HYDROXY-3-METHYLGLUTARYL COENZYME A REDUCTASE.
GN    HMGS.
OS    Streptomyces sp.
OC    Bacteria; Firmicutes; Actinobacteria; Actinobacter
OC    Actinomycetales, Streptomycineae, Streptomycetace
RN    [1]
RP    SEQUENCE FROM N.A.
RC    STRAIN=KO-3988;
RA    DAIRI T., MOTOHIRA Y., KUSUYAMA I., TAKAHASHI S.,
RT    "Cloning of the Gene Encoding 3-Hydroxy-3-Meth,lu
RT    Reductase from Terpenoid Antibiotics-Producing Str
RT    Strains.";
RL    Submitted (DEC-1998) to the EMBL/GenBank/DDBJ data
DR    EMBL; AB015627; BAA74566.1; -.
SQ    SEQUENCE   353 AA;  37958 MW;  F70DDFDC CRC32;

  Query Match           14.0%;  Score 89;  DB 2;  Len
  Best Local Similarity  31.8%;  Pred. No. 3.46e-01;
  Matches   14;  Conservative    9;  Mismatches  20;

Db      77 DAQTAYTAARAIEARIEELFAVYK-I-SKPAELLQVRHEIASNL
           |   ||| |  |  | . :  : :: :| |  |  :::
Qy      35 DVNNIYTANRNEFITTEEYKVFVNEACHPYPVILPDPSVLSGDF


RESULT    9
ID    Q19045       PRELIMINARY;       PRT;  875 AA.
AC    Q19045;
DT    01-NOV-1996 (TrEMBLrel. 01, Created)
DT    01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT    01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE    E03A3.6 PROTEIN.
GN    E03A3.6.
OS    Caenorhabditis elegans.
OC    Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdi
OC    Rhabditina; Rhabditoidea; Rhabditidae; Peloderina
RN    [1]
RP    SEQUENCE FROM N.A.
RA    GARDNER A.;
RL    Submitted (OCT-1994) to the EMBL/GenBank/DDBJ dat
RN    [2]
RP    SEQUENCE FROM N.A.
RX    MEDLINE; 94150718.
RA    WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C.,
RA    BONFIELD J., BURTON J., CONNELL M., COPSEY T., Cu
RA    CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A.
RA    GARDNER A., GREEN P., HAWKINS T., HILLIER L., Jin
RA    JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LAI
RA    LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B.
RA    PARSONS J., PERCY C., RIPKEN L., ROOPRA A., SAUNI
RA    SMALDON N., SMITH A., SONNHAMMER E., STADEN R., S
RA    THIERRY MIEG J., THIERRY MIEG D., VAUDIN M., VAUGHAN K.
RA    WATSON A., WATSON D., WATERSTON R., WOHLDMAN P., WOR
RT    "2.2 Mb of contiguous nucleotide sequence from ch
RT    elegans.";
RL    Nature 368:32-38(1994).
DR    EMBL; Z38112; CAA86231.1;
SQ    SEQUENCE   875 AA;  98924 MW;  E62D4DCE CRC32;

  Query Match           13.9%;  Score 88;  DB 5;  Len
  Best Local Similarity  23.1%;  Pred. No. 4.91e-01;
  Matches   15;  Conservative   19;  Mismatches  27;

Db     538 RIKEGLGRLLALMPYLVISLEIWSR-VMPKWLEAIVND-CIEDN
           ::   :: |:|| :|   : :::   :: :|  :: ||:  | :
Qy       2 KVQATFATLIALAAF-IFAASAKKE-IQSYKDGDVNNIYIANR
```

```
Db     596 CPLPF 600
           |  |:
Qy      61 CH-PY 64


RESULT   10
ID   O72729       PRELIMINARY;       PRT;   219 AA.
AC   O72729;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   D6L PROTEIN.
GN   D6L.
OS   Cowpox virus (CPV).
OC   Viruses, dsDNA viruses, no RNA stage, Poxviridae, Chordopoxvirinae,
OC   Orthopoxvirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GRI-90;
RX   MEDLINE; 98229462.
RA   SHCHELKUNOV S.N., SAFRONOV P.F., TOTMENIN A.V., PETROV N.A.,
RA   RYAZANKINA O.I., GUTOROV V.V., KOTWAL G.J.;
RT   "The genomic sequence analysis of the left and right species specific
RT   terminal region of a cowpox virus strain reveals unique sequences and
RT   a cluster of intact ORFs for immunomodulatory and host range
RT   proteins.";
RL   Virology 243:432-460(1998).
DR   EMBL; Y11842; CAA72582.1; -.
SQ   SEQUENCE   219 AA;  25059 MW;  6464830B CRC32;

  Query Match           13.7%;  Score 87;  DB 14;  Length 219;
  Best Local Similarity  25.8%;  Pred. No. 6.93e-01;
  Matches   16;  Conservative   15;  Mismatches  29;  Indels   2;  Gaps   2;

Db      84 IVRNLNIDDMGPVEKEFTIYEDLAIEKYGKPFIRDACVTVLDVQFDRDAFDSDSDSGSDM 143
           |:.  :  |:  ::      |:::|| | | |: :||     :  ||  : :|  |:
Qy      27 IIQRYKDGDVNNIYIA-NRNEEITIEEY-KVFVNEACHPYPVILPDRSVLSGDFTSAYAD 84

Db     144 DD 145
           ||
Qy      85 DD 86


RESULT   11
ID   O57964       PRELIMINARY;       PRT;   301 AA.
AC   O57964;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   301AA LONG HYPOTHETICAL PROTEIN.
GN   PH0225.
OS   Pyrococcus horikoshii.
OC   Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OT3;
RX   MEDLINE; 98344137.
RA   KAWARABAYASI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
RA   YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOYAMA A., NAGAI Y.,
RA   SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,
RA   FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,
RA   AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,
RA   KIKUCHI H.;
RT   "Complete sequence and gene organization of the genome of a hyper-
RT   thermophilic archaebacterium, Pyrococcus horikoshii OT3.";
RL   DNA Res. 5:55-76(1998).
DR   EMBL; AP000001; BAA29297.1; -.
SQ   SEQUENCE   301 AA;  33377 MW;  16C10EBD CRC32;

  Query Match           13.7%;  Score 87;  DB 1;  Length 301;
  Best Local Similarity  34.5%;  Pred. No. 6.93e-01;
  Matches   10;  Conservative   10;  Mismatches   8;  Indels   1;  Gaps   1;

Db      45 IVERFK-GDTFRVFAVSYDERNTIDKYKI 72
           |:::|:| ||   ::: : :|  ||: ||:
Qy      27 IIQRYKDGDVNNIYTANRNEEITIEEYKV 55


RESULT   12
ID   Q9Y8C9       PRELIMINARY;       PRT;   326 AA.
AC   Q9Y8C9;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   MATING TYPE PROTEIN MAT-2.
GN   MAT-2.
OS   Cochliobolus cymbopogonis.
OC   Eukaryota, Fungi, Ascomycota, Euascomycetes, Loculoascomycetes;
OC   Pleosporales, Pleosporaceae, Cochliobolus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. 88109-1;
RX   MEDLINE, 99254086.
RA   YUN S.H., BERBEE M.L., YODER O.C., TURGEON B.G.;
RT   "Evolution of the fungal self-fertile reproductive life style from
RT   self sterile ancestors.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:5592-5597(1999).
DR   EMBL; AF129745; AAD33447.1; -.
SQ   SEQUENCE   326 AA;  35924 MW;  254797CE CRC32;

  Query Match           13.7%;  Score 87;  DB 3;  Length 326;
  Best Local Similarity  38.7%;  Pred. No. 6.93e-01;
  Matches   12;  Conservative   7;  Mismatches  11;  Indels   1;  Gaps   1;

Db      59 ILVEHFKRFVGEAC-AFPVNLIAMDGGNGNY 88
           | .| .| || ||| :.|| |      :|::
Qy      48 ITIEEYKVFVNEACHPYPVILPDRSVLSGDF 78


RESULT   13
ID   O58489       PRELIMINARY;       PRT;   391 AA.
AC   O58489;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   391AA LONG HYPOTHETICAL ASPARTATE AMINOTRANSFERASE.
GN   PH0771.
OS   Pyrococcus horikoshii.
OC   Archaea, Euryarchaeota, Thermococcales; Thermococcaceae; Pyrococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OT3;
RX   MEDLINE; 98344137.
RA   KAWARABAYASI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
RA   YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOYAMA A., NAGAI Y.,
RA   SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,
RA   FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,
RA   AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,
RA   KIKUCHI H.;
RT   "Complete sequence and gene organization of the genome of a hyper
RT   thermophilic archaebacterium, Pyrococcus horikoshii OT3.";
RL   DNA Res. 5:55-76(1998).
DR   EMBL; AP000003; BAA29863.1; -.
DR   PFAM; PF00155; aminotran_1; 1.
KW   Transferase, Aminotransferase.
SQ   SEQUENCE   391 AA;  44565 MW;  A549ACDA CRC32;

  Query Match           13.7%;  Score 87;  DB 1;  Length 391;
  Best Local Similarity  31.4%;  Pred. No. 6.93e-01;
  Matches   11;  Conservative   13;  Mismatches  12;  Indels   2;  Gaps   2;

Db     205 EPYEHFIYEDAKHYPMIKFAPENTILANSFSKTFA 239
           | |  |: |    ||:| : |: ::|:  |: ::|
Qy      51 EEYKVFVNEACHPYPVI-L PDRSVLSGDFTSAYA 83
```

```
RESULT   14
ID   Q9ZPS7      PRELIMINARY;      PRT:   592 AA.
AC   Q9ZPS7;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   PUTATIVE MULTISPANNING MEMBRANE PROTEIN.
GN   F14H20.4.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC   Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   LIN X., KAUL S., SHEA T.P., FUJII C.Y., SHEN M., VANAKEN S.E.,
RA   BARNSTEAD M.E., MASON T.M., BOWMAN C.L., RONNING C.M., BENITO M.,
RA   CARRERA A.J., CREASY T.H., BUELL C.R., TOWN C.D., NIERMAN W.C.,
RA   FRASER C.M., VENTER J.C.;
RT   "Arabidopsis thaliana chromosome II BAC F14H20 genomic sequence.";
RL   Submitted (MAR 1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC006532; AAD20090.1; -.
SQ   SEQUENCE   592 AA;  68048 MW;  AD675F42 CRC32;

  Query Match             13.7%;  Score 87;  DB 10;  Length 592;
  Best Local Similarity   26.4%;  Pred. No. 6.93e-01;
  Matches   15;  Conservative   14;  Mismatches  27;  Indels   1;  Gaps   1;

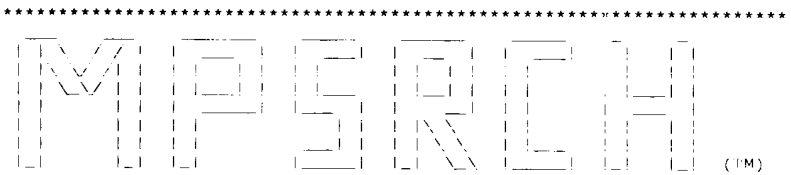
Db       1 MRLPITLLLFICALIFSGAGIVKSDASDHRYKDGDSVPLYANKVGPFHNPSETYKYF 57
           |:: :|:  :|:   :  |:  : |   :||||||   :|: :        | |: |
Qy       1 MKVQATFATIIALAAYFPAANAWFSVIIQPYKDSDVNNIYTANENE-EITIEEYKVF 56


RESULT   15
ID   Q51162      PRELIMINARY;      PRT:   720 AA.
AC   Q51162;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   FE-REGULATED PROTEIN B PRECURSOR.
GN   FRPB.
OS   Neisseria meningitidis.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B44/76;
RX   MEDLINE; 96009786.
RA   PETTERSSON A., MAAS A., VAN WASSENAAR D., VAN DER LEY P, TOMMASSEN J.;
RT   "Molecular characterization of FrpB, the 70-kilodalton iron-regulated
RT   outer membrane protein of Neisseria meningitidis.";
RL   Infect. Immun. 63:4181-4184(1995).
DR   EMBL; X89755; CAA61902.1; -.
KW   Signal.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    720
SQ   SEQUENCE   720 AA;  79611 MW;  4F79FA16 CRC32;

  Query Match             13.7%;  Score 87;  DB 2;  Length 720;
  Best Local Similarity   34.5%;  Pred. No. 6.93e-01;
  Matches   20;  Conservative   15;  Mismatches  18;  Indels   5;  Gaps   4;

Db     186 KEGNFDGLFSYNRNDFKDYFAGKGFRNVNGGKTVPYSA-LDKPSYLA-KIGTTFGDDD 241
           |:|: : ::: |||:|   |  | |  || :  ||: | || |:  :  :::|||
Qy      32 KDGDVNNIYTANRNEEITIEEYKVF--VNEA-CHPYPVILPDRSVLSGDFISAYADDD 86


Search completed: Fri Aug 25 17:54:52 2000
Job time : 64 secs.
```



Release 4.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U K
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 25 17:52:27 2000; MasPar time 7.26 Seconds
578.086 Million cell updates/sec

Tabular output not generated.

Title: >US-09-257-585-4
Description: (1-89) from US09257585.pep
Perfect Score: 634
Sequence: 1 MKVQAIFAILIALAAYFPAA.....DRSVLSGDFISAYADDDESC 89

Scoring table: PAM 150
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir64
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 39.486; Variance 73.326; scale 0.539

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	93	14.7	191	1	H64744	yaeD protein - Escher	2.13e-01
2	93	14.7	1191	2	A53491	bumetanide-sensitive	2.13e-01
3	93	14.7	1205	2	A55015	bumetanide sensitive	2.13e-01
4	93	14.7	1212	2	A57187	bumetanide-sensitive	2.13e-01
5	92	14.5	637	2	S42251	probable nucleotide t	2.95e-01
6	90	14.2	1526	2	A45605	mature parasite infec	5.62e-01
7	89	14.0	427	2	A60048	cyclin B3 - yeast (Sa	7.74e-01
8	88	13.9	126	2	S11782	regulatory protein bl	1.06e+00
9	87	13.7	301	2	H71246	hypothetical protein	1.45e+00
10	87	13.7	391	2	E71125	probable aspartate am	1.45e+00
11	87	13.7	720	2	S59133	Fe-regulated protein	1.45e+00
12	87	13.7	922	2	D70066	SNF2 helicase homolog	1.45e+00
13	86	13.6	192	2	D71211	hypothetical protein	1.99e+00
14	86	13.6	4951	1	VFIHB1	F1 protein - avian in	1.99e+00
15	85	13.4	260	2	S01506	cytochrome c oxidase	2.71e+00
16	85	13.4	492	2	S71245	glucose 6 phosphate 1	2.71e+00
17	85	13.4	773	2	I59351	carnitine O-palmitoyl	2.71e+00
18	84	13.2	190	1	KLSSAA	calcium-binding prote	3.68e+00
19	84	13.2	739	2	I15215	hypothetical protein	3.68e+00
20	84	13.2	1252	2	S52004	phospholipase C (EC 3	3.68e+00
21	83	13.1	376	2	S67085	hypothetical protein	4.99e+00
22	82	12.9	192	1	KLSSAB	calcium-binding prote	6.75e+00
23	82	12.9	286	2	JQ2289	floral homeotic prote	6.75e+00
24	82	12.9	342	2	T37447	hypothetica	6.75e+00
25	82	12.9	458	2	T16214	hypothetica	6.75e+00
26	82	12.9	617	2	A71194	hypothetica	6.75e+00
27	82	12.9	734	2	A57166	acyl-CoA ox	6.75e+00
28	82	12.9	1118	1	SYBYPI	carbamoyl-p	6.75e+00
29	82	12.9	1249	2	S54479	tripeptidy	6.75e+00
30	81	12.8	125	2	S41190	ribosomal p	8.80e+00
31	81	12.8	134	2	I50299	TIAda.33	8.80e+00
32	81	12.8	246	2	A70418	conserved h	8.80e+00
33	81	12.8	350	2	S25087	aspartate	8.80e+00
34	81	12.8	401	2	H69833	conserved h	8.80e+00
35	81	12.8	1172	2	S21109	hypothetica	8.80e+00
36	81	12.8	1449	2	S47429	E2 glycopr	8.80e+00
37	81	12.8	1449	2	A45929	E2 glycopr	8.80e+00
38	81	12.8	1449	3	VCIBB5	E2 glycopr	8.80e+00
39	80	12.6	206	2	S29495	hypothetica	1.15e+01
40	80	12.6	441	2	A48455	acidic pho	1.15e+01
41	80	12.6	463	2	A28826	regulatory	1.15e+01
42	80	12.6	479	2	I61922	hypothetica	1.15e+01
43	80	12.6	525	2	I05109	hypothetica	1.15e+01
44	80	12.6	679	2	S73451	hypothetica	1.15e+01
45	80	12.6	1040	2	I60156	CoM1 like p	1.15e+01

ALIGNMENTS

RESULT 1
ENTRY H64744 #type complete
TITLE yaeD protein - Escherichia coli
ORGANISM #formal_name Escherichia coli
DATE 10-Sep-1999 #sequence_revision 10-Sep 10-Sep 1999
ACCESSIONS H64744
REFERENCE A64720
#authors Blattner, F.R.; Plunkett III, G.; Bloc Burland, V.; Riley, M.; Collado Vid Rode, C.K.; Mayhew, G.F.; Gregor, J Kirkpatrick, H.A.; Goeden, M.A.; Ro y.
#journal Science (1997) 277:1453-1462
#title The complete genome sequence of Escher
#cross-references MUID:97426617
#accession H64744
##status nucleic acid sequence not shown;
##molecule_type DNA
##residues 1-191 ##label BLAT
##cross-references GB:AE000129; GB:U00096; NID DBGP:b0200
##experimental_source strain K 12, substrain MG1
GENETICS
#gene yaeD
CLASSIFICATION #superfamily ORF protein
SUMMARY #length 191 #molecular-weight 21294

Query Match 14.7%; Score 93; DB 1; Len
Best Local Similarity 26.5%; Pred. No. 2.13e-
Matches 18; Conservative 21; Mismatches 24;

Db 68 FETLIEWMDWSLAD RFVDLLALYYCPHRPQGSVEEFRQLVCD
Qy 17 FPAANAWEPCILEPE SYNRSTTCNENLEIIELTAVPYNEA

Db 125 HIDMAASY 132
Qy 75 SGDRSAY 82

RESULT 2
ENTRY A53491 #type complete
TITLE bumetanide-sensitive Na-K-Cl cotransp
ORGANISM #formal_name Squalus acanthias #comme
DATE 02-Jun-1994 #sequence_revision 02-Jun 17-Mar-1999

```
ACCESSIONS      A53491
REFERENCE       A53491
  #authors      Xu, J.C.; Lytle, C.; Zhu, T.T.; Payne, J.A.; Benz Jr., E.;
                Forbush III, B.
  #journal      Proc. Natl. Acad. Sci. U.S.A. (1994) 91:2201-2205
  #title        Molecular cloning and functional expression of the
                bumetanide-sensitive Na-K-Cl cotransporter
  #cross-references MUID:94181560
  #accession    A53491
    ##status        preliminary
    ##molecule_type mRNA
    ##residues      1-1191 ##label XUA
    ##cross-references GB:U05958; NID:g454096; PID:g454097
SUMMARY         #length 1191  #molecular-weight 129774  #checksum 7133

  Query Match           14.7%;  Score 93;  DB 2;  Length 1191;
  Best Local Similarity  25.5%;  Pred. No. 2.13e-01;
  Matches    13;  Conservative   13;  Mismatches  23;  Indels   2;  Gaps   2;

Db   1014 LTLLIPYLLITKKKWKDCKIRVFTGGKINFIDHDRRTMATLLSKFRIDFSD 1064
            |  | :|: :::  |||| |: :  | :| |    |     :  ::: | :
Qy     10 LIALAAYFPAANA-WKDCIIQRYKDGDVNNIYTANRNEEITEEYKV-FVN 58


RESULT    3
ENTRY           A55015      #type complete
TITLE           bumetanide-sensitive Na(+)-K(+)-2Cl- cotransporter protein
                mouse
ORGANISM        #formal_name Mus musculus #common_name house mouse
DATE            11-Nov-1994 #sequence_revision 11-Nov-1994 #text_change
                07-May-1999
ACCESSIONS      A55015
REFERENCE       A55015
  #authors      Delpire, E.; Rauchman, M.I.; Beier, D.R.; Hebert, S.C.;
                Gullans, S.R.
  #journal      J. Biol. Chem. (1994) 269:25677-25683
  #title        'Molecular cloning and chromosome localization of a putative
                basolateral Na(+)-K(+)-2Cl- cotransporter from mouse inner
                medullary collecting duct (mIMCD-3) cells.
  #cross-references MUID:95014474
  #accession    A55015
    ##status        preliminary
    ##molecule_type mRNA
    ##residues      1-1205 ##label DEL
    ##cross-references GB:U13174; NID:g560007; PID:g560008
SUMMARY         #length 1205  #molecular-weight 130949  #checksum 6188

  Query Match           14.7%;  Score 93;  DB 2;  Length 1205;
  Best Local Similarity  25.5%;  Pred. No. 2.13e-01;
  Matches    13;  Conservative   13;  Mismatches  23;  Indels   2;  Gaps   2;

Db   1028 LTLLIPYLLITKKKWKDCKIRVFTGGKINFIDHDRRAMATLLSKFRIDFSD 1078
            |  | :|: :::  |||| |: :  | :| |    |     :  ::: | :
Qy     10 LIALAAYFPAANA-WKDCIIQRYKDGDVNNIYTANRNEEITEEYKV-FVN 58


RESULT    4
ENTRY           A57187      #type complete
TITLE           bumetanide-sensitive Na-K-Cl cotransporter - human
ORGANISM        #formal_name Homo sapiens #common_name man
DATE            16-Feb-1996 #sequence_revision 16-Feb-1996 #text_change
                28-Feb-1997
ACCESSIONS      A57187
REFERENCE       A57187
  #authors      Payne, J.A.; Xu, J.C.; Haas, M.; Lytle, C.Y.; Ward, D.;
                Forbush, B.
  #journal      J. Biol. Chem. (1995) 270:17977-17985
  #title        Primary structure, functional expression, and chromosomal
                localization of the bumetanide-sensitive Na-K-Cl
                cotransporter in human colon.
  #cross-references MUID:95355397
  #accession    A57187
    ##status        preliminary
    ##molecule_type mRNA
    ##residues      1-1212 ##label RES
    ##cross-references EMBL:S30246, NID:g300681, PID:g300682
GENETICS
  #gene         NKCC1
  #map_position 5q23.3
SUMMARY         #length 1212  #molecular-weight 131446  #checksum 2210

  Query Match           14.7%;  Score 93;  DB 2;  Length 1212;
  Best Local Similarity  25.5%;  Pred. No. 2.13e-01;
  Matches    13;  Conservative   13;  Mismatches  23;  Indels   2;  Gaps   2;

Db   1035 LTLLIPYLLITKKKWKDCKIRVFTGGKINFIDHDRRAMATLLSKFRIDFSD 1085
            |  | :|: :::  |||| |: :  | :| |    |     :  ::: | :
Qy     10 LIALAAYFPAANA-WKDCIIQRYKDGDVNNIYTANRNEEITEEYKV-FVN 58


RESULT    5
ENTRY           S42251      #type complete
TITLE           probable nucleoside triphosphatase (EC 3.6.1.15) I - fowlpox
                virus
ORGANISM        #formal_name fowlpox virus
DATE            19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change
                07-May-1999
ACCESSIONS      S42251
REFERENCE       S42247
  #authors      Binns, M.M.; Britton, B.S.; Mason, C.; Boursnell, M.E.G.
  #journal      J. Gen. Virol. (1990) 71:2873-2881
  #title        Analysis of the fowlpox virus genome region corresponding to
                the vaccinia virus D5 to A1 region: location of, and
                variation in, non-essential genes in poxviruses.
  #cross-references MUID:91108382
  #accession    S42251
    ##molecule_type DNA
    ##residues      1-637 ##label BIN
    ##experimental_source strain HP444
CLASSIFICATION  #superfamily vaccinia virus nucleoside-triphosphatase I;
                DEAD/H box helicase homology
KEYWORDS        hydrolase
FEATURE
  56-481            #domain DEAD/H box helicase homology #label DEAD\
  142-145           #region DEXH motif
SUMMARY         #length 637  #molecular-weight 73774  #checksum 4242

  Query Match           14.5%;  Score 92;  DB 2;  Length 637;
  Best Local Similarity  31.0%;  Pred. No. 2.95e-01;
  Matches    13;  Conservative   11;  Mismatches  15;  Indels   3;  Gaps   3;

Db    108 KDCIIMNYDDQNFHNKFFTNIKS-INVKS-RIFIIIDECHNF 147
          |||||  | | : :| : :| :  | :   ::|:  : || :
Qy     24 KDCIIQRYKDGDVNNIYTANRNEEITEEYKVFVN-EACHPY 64


RESULT    6
ENTRY           A45605      #type complete
TITLE           mature-parasite-infected erythrocyte surface antigen MESA -
                Plasmodium falciparum
ORGANISM        #formal_name Plasmodium falciparum
DATE            22-Apr-1993 #sequence_revision 18 Nov 1994 #text_change
                07-Feb-1997
ACCESSIONS      A45605
REFERENCE       A45605
  #authors      Coppel, R.L.
  #journal      Mol. Biochem. Parasitol. (1992) 50:335-347
  #title        Repeat structures in a Plasmodium falciparum protein (MESA)
                that binds human erythrocyte protein 4.1.
  #cross-references MUID:92158014
  #accession    A45605
    ##status        preliminary
    ##molecule_type DNA
    ##residues      1-1526 ##label COP
    ##experimental_source Papua New Guinean isolate FC27
    ##note          sequence extracted from NCBI backbone (NCBIN:83648,
```

NCBIP:83656)
KEYWORDS surface antigen
SUMMARY #length 1526 #molecular weight 179141 #checksum 6526

Query Match 14.2%; Score 90; DB 2; Length 1526;
Best Local Similarity 24.6%; Pred. No. 5.62e-01;
Matches 15; Conservative 21; Mismatches 21; Indels 4; Gaps 4;

Db 1373 RRNCKE-NEYNNG[illegible] 1429
Qy 23 WKIK[illegible] 81

Db 1430 F 1430
Qy 82 Y 82

RESULT 7
ENTRY A60048 #type complete
TITLE cyclin B4 - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES protein D1539; protein YDL155w
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Jul-1999
ACCESSIONS A60048; A46228; S14167; S67703; S31291
REFERENCE A60048
#authors Fitch, I.; Dahmann, C.; Surana, U.; Amon, A.; Nasmyth, K.; Goetsch, L.; Byers, B.; Futcher, B.
#journal Mol. Biol. Cell (1992) 3:805-818
#title Characterization of four B-type cyclin genes of the budding yeast Saccharomyces cerevisiae.
#cross-references MUID:92385891
#accession A60048
##molecule_type DNA
##residues 1-427 ##label FIT
##cross-references GB:M80302; NID:g171911; PIDN:AAA34765.1; PID:g171912
REFERENCE A46228
#authors Richardson, H.; Lew, D.J.; Henze, M.; Sugimoto, K.; Reed, S.I.
#journal Genes Dev. (1992) 6:2021-2034
#title Cyclin-B homologs in Saccharomyces cerevisiae function in S phase and in G2.
#cross-references MUID:93051323
#accession A46228
##molecule_type DNA
##residues 1-47,'R',49-427 ##label KIC
##cross-references EMBL:X69425; NID:g5524; PIDN:CAA49101.1; PID:g5525
##note sequence extracted from NCBI backbone (NCBIN:117852, NCBIP:117853)
REFERENCE S14165
#authors Surana, U.; Robitsch, H.; Price, C.; Schuster, T.; Fitch, I.; Futcher, A.B.; Nasmyth, K.
#journal Cell (1991) 65:145-161
#title The role of CDC28 and cyclins during mitosis in the budding yeast S. cerevisiae.
#cross-references MUID:91191554
#accession S14167
##status not compared with conceptual translation
##molecule_type DNA
##residues 194-383 ##label SUR
REFERENCE S67693
#authors Perea, J.; Bluqeon, C.; Delaveau, T.; Jacq, C.
#submission submitted to the Protein Sequence Database, July 1996
#accession S67703
##molecule_type DNA
##residues 1-427 ##label PER
##cross-references EMBL:274203; NID:g1431244; PIDN:CAA98729.1; PID:g253065; PID:g1431245; MIPS:YDL155w
##experimental_source strain S288C
GENETICS
#gene SGD:CLB3
##cross-references SGD:S0002314; MIPS:YDL155w
#map_position 4L

CLASSIFICATION #superfamily cyclin
KEYWORDS cell cycle control
SUMMARY #length 427 #molecular weight 49471 #

Query Match 14.0%; Score 89; DB 2; Length
Best Local Similarity 24.6%; Pred. No. 7.74e-01;
Matches 15; Conservative 22; Mismatches 17;

Db 265 [illegible] 321
Qy 24 [illegible] 80

Db 294 F 294
Qy 78 F 78

RESULT 8
ENTRY S11782 #type complete
TITLE regulatory protein blaI - Staphylococ
ORGANISM #formal_name Staphylococcus aureus
DATE 03-May-1994 #sequence_revision 26-Feb- 26-Aug-1999
ACCESSIONS S11782; S34446
REFERENCE S11779
#authors Rowland, S.J.; Dyke, K.G.H.
#journal Mol. Microbiol. (1990) 4:961-975
#title Tn552, a novel transposable element fr aureus.
#cross-references MUID:91014666
#accession S11782
##molecule_type DNA
##residues 1-12[illegible] ##label MOW
##cross-references EMBL:X52734; NID:g46754; PIDN
##experimental_source transposon Tn552
##note the authors translated the codon Trp
REFERENCE S34444
#authors Wang, P.Z.; Projan, S.J.; Novick, R.P
#journal Nucleic Acids Res. (1991) 19:4066
#title Nucleotide sequence of beta-lactamase Staphylococcal plasmid p1258.
#cross-references MUID:91319662
#accession S34446
##status nucleic acid sequence not st
##molecule_type DNA
##residues 1-12[illegible] ##label WAN
##cross-references EMBL:M62456; NID:g152964; PIP PID:g152967
##experimental_source plasmid p1258
##note the nucleotide sequence was s library September 199
GENETICS
#gene blaI
#genome plasmid
CLASSIFICATION #superfamily regulatory protein blaI
KEYWORDS DNA binding; transcription regulati
SUMMARY #length 126 #molecular-weight 14876

Query Match 13.9%; Score 88; DB 2; Le
Best Local Similarity 33.3%; Pred. No. 1.06e-
Matches 12; Conservative 12; Mismatches 9;

Db 55 [illegible] 110
Qy 24 [illegible] 79

RESULT 9
ENTRY B71246 #type complete
TITLE hypothetical protein PH0225 - Pyroc
ORGANISM #formal_name Pyrococcus horikoshii
DATE 14-Aug-1998 #sequence_revision 14-A

```
                20-Sep-1999
ACCESSIONS      B71246
REFERENCE       A71000
   #authors     Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.;
                  Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.;
                  Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.;
                  Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.;
                  Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi,
                  A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.;
                  Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.
   #journal     DNA Res. (1998) 5:55-76
   #title       Complete sequence and gene organization of the genome of a
                  hyper-thermophilic archaebacterium, Pyrococcus horikoshii
                  OT3.
   #cross-references MUID:98344137
   #accession   B71246
      ##status       preliminary; nucleic acid sequence not shown;
                       translation not shown
      ##molecule_type DNA
      ##residues     1-301 ##label KAW
      ##cross-references GB:AP000001; NID:g3236128; PID:d1030210; PID:g3256614
      ##experimental_source strain OT3
      ##note         this accession replaces an interim accession for a
                       sequence replaced by GenBank
GENETICS
   #gene        PH0225
CLASSIFICATION  #superfamily Pyrococcus horikoshii hypothetical protein
                  PH0225
SUMMARY         #length 301  #molecular-weight 33377  #checksum 9682

  Query Match           13.7%;  Score 87;  DB 2;  Length 301;
  Best Local Similarity  34.5%;  Pred. No. 1.45e+00;
  Matches    10;  Conservative   10;  Mismatches   8;  Indels   1;  Gaps   1;

Db      45 IVERFK-GDTFRVFAVSYDEGNTIDKYK1 72
           |:::|:| ||   ::: : :|  ||: ||:
Qy      27 IIQRYKDGDVNNIYTANENEEITIEEYKV 55


RESULT   10
ENTRY           E71125      #type complete
TITLE           probable aspartate aminotransferase - Pyrococcus horikoshii
ORGANISM        #formal_name Pyrococcus horikoshii
DATE            14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change
                  26-Aug-1999
ACCESSIONS      E71125
REFERENCE       A71000
   #authors     Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.;
                  Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.;
                  Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.;
                  Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.;
                  Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi,
                  A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.;
                  Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.
   #journal     DNA Res. (1998) 5:55-76
   #title       Complete sequence and gene organization of the genome of a
                  hyper-thermophilic archaebacterium, Pyrococcus horikoshii
                  OT3.
   #cross-references MUID:98344137
   #accession   E71125
      ##status       preliminary; nucleic acid sequence not shown;
                       translation not shown
      ##molecule_type DNA
      ##residues     1-491 ##label KAW
      ##cross-references GB:AP000003; NID:g3236130; PID:d1030806; PID:g3257180
      ##experimental_source strain OT3
      ##note         this accession replaces an interim accession for a
                       sequence replaced by GenBank
GENETICS
   #gene        PH0771
CLASSIFICATION  #superfamily aspartate transaminase
SUMMARY         #length 491  #molecular-weight 44565  #checksum 193

  Query Match           13.7%;  Score 87;  DB 2;  Length 491;
  Best Local Similarity  31.4%;  Pred. No. 1.45e+00;
  Matches    11;  Conservative   10;  Mismatches  12;  Indels   2;  Gaps   2;

Db     205 FDVFRHFIYFDAKHYPMIKEAPENFIIANSESKIFA 239
           | |  |: |    ||:| : |: ::|:  |: ::|
Qy      51 EEYKVEVNEAGHPYPVI-L-PDRSVLSGDEISAYA 83


RESULT   11
ENTRY           S58133      #type complete
TITLE           Fe-regulated protein B precursor - Neisseria meningitidis
ORGANISM        #formal_name Neisseria meningitidis
DATE            13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change
                  29-Sep-1999
ACCESSIONS      S58133
REFERENCE       S58133
   #authors     Pettersson, A.M.; Maas, X.Y.Z., van Wassenaar, X.Y.Z., van
                  der Ley, P.; Tommassen, X.Y.Z.
   #submission  submitted to the EMBL Data Library, July 1995
   #description Molecular characterization of FrpB, the 70 kilodalton
                  iron-regulated outer membrane protein of Neisseria
                  meningitidis.
   #accession   S58133
      ##status       preliminary
      ##molecule_type DNA
      ##residues     1-720 ##label PET
      ##cross-references EMBL:X89755; NID:g1052769; PID:g1052770
CLASSIFICATION  #superfamily tonB-dependent receptor amino-terminal homology;
                  tonB-dependent receptor carboxyl-terminal homology
FEATURE
   62 191           #domain tonB dependent receptor amino-terminal homology
                      #label TNN
SUMMARY         #length 720  #molecular-weight 79611  #checksum 8332

  Query Match           13.7%;  Score 87;  DB 2;  Length 720;
  Best Local Similarity  34.5%;  Pred. No. 1.45e+00;
  Matches    20;  Conservative   15;  Mismatches  18;  Indels   5;  Gaps   4;

Db     186 KEGNFDLELSYNKNDEKDYEAGNGFKNVNGGNIVRYSA-EDKRSYLA-KIGTIFGDDD 241
           |:|: : ::: |||:|   | | |  || :   ||:  |  || |:  : ::::|||
Qy      47 KTGTVNNIYTANENEEIIIEYFVE  VNEA-CHEYFVILPDRSVLSGDEISAYADDD 86


RESULT   12
ENTRY           D70066      #type complete
TITLE           SNF2 helicase homolog ywqA - Bacillus subtilis
ORGANISM        #formal_name Bacillus subtilis
DATE            05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
                  24-Sep-1998
ACCESSIONS      D70066
REFERENCE       A69580
   #authors     Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
                  Alloni, G., Azevedo, V., Bertero, M.G., Bessieres, P.;
                  Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
                  A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
                  Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
                  Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
                  Daniel, R.A., Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
                  Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.;
                  Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
                  M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
                  S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
                  Guiseppi, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood,
                  C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
                  Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
                  Kasahara, Y.; Klaerr Blanchard, M.; Klein, C.; Kobayashi,
                  Y., Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.;
                  Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
                  Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
                  Mauel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
                  M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly,
                  M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
```

```
                  V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott,
                  A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
                  Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
                  Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
                  Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
                  Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
                  B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
                  Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
                  Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
                  Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
                  Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.;
                  Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
                  K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
                  Yoshikawa, H.; Danchin, A.
   #journal       Nature (1997) 390:249-256
   #title         The complete genome sequence of the Gram-positive bacterium
                  Bacillus subtilis.
   #cross-references MUID:98044033
   #accession     D70666
      ##status          preliminary; nucleic acid sequence not shown;
                         translation not shown
      ##molecule_type DNA
      ##residues       1-922 ##label KUN
      ##cross-references GB:Z99122; GB:AL009126; NID:g2633029; PID:e1184534;
                         PID:g2636153
      ##experimental_source strain 168
GENETICS
   #gene          ywqA
SUMMARY           #length 922  #molecular-weight 106029  #checksum 1993

  Query Match             13.7%;  Score 87;  DB 2;  Length 922;
  Best Local Similarity  19.0%;  Pred. No. 1.45e+00;
  Matches    16;  Conservative   29;  Mismatches  35;  Indels    4;  Gaps    4;

Db     100 STIYEFIADGDFMPDYDAWTNGVF-RWKDRDNILEGFTAEWFSA-AVQDYIQYDDDLREK 157
           :|:  :|| : : |  :|| : :: | ||  :: : :|||:    ::::|  | ::
Qy       5 AIFAILIALAAYFPAANAWKICIIQRYKD-CDVNNIYIANKNEKIIEEYKVPVNEACHF 63

Db     158 WEHI-KEKSPAVTIFRGHFLDEED 180
           |  ::|    | : : |:::
Qy      64 YPVILPDRSVLSGDFTSAYADDDE 87


RESULT   13
ENTRY             D71211      #type complete
TITLE             hypothetical protein PH1956 - Pyrococcus horikoshii
ORGANISM          #formal_name Pyrococcus horikoshii
DATE              14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change
                    14-Aug-1998
ACCESSIONS        D71211
REFERENCE         A71000
   #authors       Kawarabayasi, Y.; Sawada, M.; Horikawa, H.  Haikawa, Y.;
                    Hino, Y.; Yamamoto, S.; Sekine, M.; Baba  S.; Kosugi, H.;
                    Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.;
                    Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.;
                    Tanaka  T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi,
                    A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.;
                    Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.
   #journal       DNA Res. (1998) 5:55-76
   #title         Complete sequence and gene organization of the genome of a
                    hyper-thermophilic archaebacterium, Pyrococcus horikoshii
                    OT3.
   #cross-references MUID:98344137
   #accession     D71211
      ##status          preliminary; nucleic acid sequence not shown;
                         translation not shown
      ##molecule_type DNA
      ##residues       1-192 ##label KAW
      ##cross-references GB:AP000007; NID:g3236134; PID:d1032026; PID:g3258400
      ##experimental_source strain OT3
      ##note           this accession replaces an interim accession for a
                         sequence replaced by GenBank
GENETICS
   #gene          PH1956
SUMMARY           #length 192  #molecular-weight 21954  #

  Query Match             1.6%;  Score 86;  DB 2;  Leng
  Best Local Similarity  24.4%;  Pred. No. 1.99e+00;
  Matches    9;  Conservative   15;  Mismatches  11;

Db      30 LSDMKCYFPLDEEIWKE LVFK L.FVVYTFVYAVEQEE 64
           |  : :||   ::|:: :::|  :  | ::|: :::|
Qy      10 LIALAAYFPAANAWKICIIQRYKD-CDVNNIYIANKNE 46


RESULT   14
ENTRY             VF1HB1      #type complete
TITLE             F1 protein  avian infectious bronchi
                    Beaudette)
ORGANISM          #formal_name avian infectious bronchi
DATE              30-Jun-1992 #sequence_revision 30-Jun
                    16-Jul-1999
ACCESSIONS        A33094
REFERENCE         A33094
   #authors       Boursnell, M.E.G.; Brown, T.D.K.; Fo
                    Tomley, F.M.; Binns, M.M.
   #journal       J. Gen. Virol. (1987) 68:57-77
   #title         Completion of the sequence of the ge
                    avian infectious bronchitis virus.
   #cross-references MUID:87113458
   #accession     A33094
      ##molecule_type genomic RNA
      ##residues       1-3951 ##label BOU
      ##cross-references GB:M21570; GB:M21486; NID:gss
                         PID:g333172
CLASSIFICATION    #superfamily avian infectious bronchi
SUMMARY           #length 3951  #molecular-weight 44111

  Query Match             1.6%;  Score 86;  DB 1;  Len
  Best Local Similarity  24.4%;  Pred. No. 1.99e+00;
  Matches    9;  Conservative   15;  Mismatches  11;

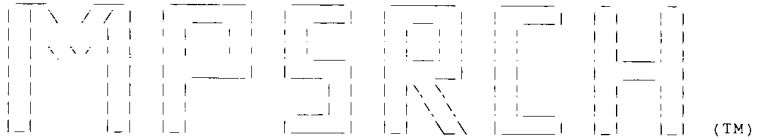
Db    2742 YLQACRAWLAYALIAAYRNSDVEIVYIPR 2770
           |: :  ||    :::|::: : :|| : :|
Qy      16 YFPAANAWKICIIQRYKDCDVNNIYIANK 44


RESULT   15
ENTRY             S01506      #type complete
TITLE             cytochrome-c oxidase (EC 1.9.3.1) cha
                    (Strongylocentrotus purpuratus) mit
ORGANISM          #formal_name mitochondrion Strongyloc
                    #common_name purple urchin
DATE              01-Dec-1989 #sequence_revision 01-Dec
                    20-Aug-1999
ACCESSIONS        S01506
REFERENCE         S01499
   #authors       Jacobs, H.T.; Elliott, D.J.; Math, V
   #journal       J. Mol. Biol. (1988) 202:185-217
   #title         Nucleotide sequence and gene organiz
                    mitochondrial DNA.
   #cross-references MUID:88301761
   #accession     S01506
      ##molecule_type DNA
      ##residues       1-26 ##label JAC
      ##cross-references EMBL:X12634;
GENETICS
   #gene          coIII
   #genome        mitochondrion
   #genetic_code SGC8
CLASSIFICATION    #superfamily cytochrome-c oxidase ch
KEYWORDS          electron transfer; membrane associat
                    inner membrane; mitochondrion; oxi
                    oxidoreductase; respiratory chain;
SUMMARY           #length 260  #molecular-weight 29489
```

Query Match 14.4%; Score 85; DB 2; Length 260;
Best Local Similarity 29.4%; Pred. No. 2.71e+00;
Matches 13; Conservative 7; Mismatches 11; Indels 2, Gaps 2;

```
Db    159 IQALFLT-VALGSYFTALQAW-EYIDAPFTIAD 189
          :|| | | :||::|| | || : |   :  :|
Qy      3 VQATFATLIALAAYFPAANAWKDCIIQRYKDGD 35
```

Search completed: Fri Aug 25 17:52:46 2000
Job time : 19 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 25 17:54:04 2000; MasPar time 4.76 Seconds
579.206 Million cell updates/sec

Tabular output not generated.

Title: >US-09-257-585-4
Description: (1-89) from US09257585.pep
Perfect Score: 634
Sequence: 1 MKVQATFATLIALAAYFPAA.....DRSVLSGDFTSAYADDDESC 89

Scoring table: PAM 150
Gap 11

Searched: 85661 seqs, 30989116 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 40.352; Variance 65.167; scale 0.619

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Prod. No.
1	93	14.7	191	1	YAED_ECOLI	HYPOTHETICAL 21.3 KDA	3.97e-02
2	93	14.7	1191	1	NKC1_SQUAC	BUMETANIDE-SENSITIVE S	3.97e-02
3	93	14.7	1205	1	NKC1_MOUSE	BUMETANIDE-SENSITIVE S	3.97e-02
4	93	14.7	1212	1	NKC1_HUMAN	BUMETANIDE-SENSITIVE S	3.97e-02
5	92	14.5	637	1	NIP1_FOWPV	NUCLEOSIDE TRIPHOSPHAT	5.73e-02
6	89	14.0	427	1	CG23_YEAST	G2/MITOTIC-SPECIFIC CY	1.70e-01
7	88	13.9	126	1	BLAI_STAAU	PENICILLINASE PEPPERSO	2.42e-01
8	86	13.6	461	1	NIFN_SYNP8	NITROGENASE IRON-MOLYB	4.89e-01
9	86	13.6	3951	1	VGF1_IBVB	F1 PROTEIN	4.89e-01
10	85	13.4	260	1	COX3_STRPU	CYTOCHROME C OXIDASE P	6.93e-01
11	85	13.4	773	1	CPT1_HUMAN	CARNITINE O-PALMITOYLT	6.93e-01
12	84	13.2	1252	1	PIP5_HUMAN	1-PHOSPHATIDYLINOSITOL	9.77e-01
13	82	12.9	192	1	SCPA_PENSP	SARCOPLASMIC CALCIUM B	1.38e+00
14	82	12.9	724	1	CAO2_CANMA	ACYL-COENZYME A OXIDAS	1.38e+00
15	82	12.9	1118	1	CARB_YEAST	CARBAMOYL-PHOSPHATE SY	1.38e+00
16	82	12.9	1249	1	TPP2_HUMAN	TRIPEPTIDYL-PEPTIDASE	1.38e+00
17	81	12.8	125	1	RT11_MARPO	MITOCHONDRIAL RIBOSOMA	2.69e+00
18	81	12.8	1447	1	VGL2_CVPPT	E2 GLYCOPROTEIN PRECUR	2.69e+00
19	81	12.8	1447	1	VGL2_CVPPR	E2 GLYCOPROTEIN PRECUR	2.69e+00
20	81	12.8	1449	1	VGL2_CVPMI	E2 GLYCOPROTEIN PRECUR	2.69e+00
21	81	12.8	1449	1	VGL2_CVPFS	E2 GLYCOPROTEIN PRECUR	2.69e+00
22	80	12.6	441	1	PHPA_PLACH	ACIDIC PHOSPHOPROTEIN	3.75e+00
23	80	12.6	463	1	KIS_DROME	DNA BINDING PROTEIN KI	3.75e+00
24	80	12.6	644	1	NUC1_MOUSE	NUCLEOLIN.	3.75e+00
25	80	12.6	715	1	VCP_BEVIB	CAPSID PROTEIN	3.75e+00
26	79	12.5	182	1	VB07_VACCV	PROTEIN B7 P	4.98e+00
27	79	12.5	749	1	SUS_HUMAN	SULFATE TRAN	4.98e+00
28	79	12.5	971	1	YLIA_YEAST	BYE PROTEIN	4.98e+00
29	79	12.5	1,035	1	NKC2_RAT	BUMETANIDE-S	4.98e+00
30	79	12.5	1099	1	NKC2_RABIT	BUMETANIDE-S	4.98e+00
31	79	12.5	1245	1	CIP5_RAT	1 PHOSPHATI	4.98e+00
32	78	12.3	144	1	GLOB_SMITH	GLOBIN PII	6.49e+00
33	78	12.3	221	1	YIEH_ECOLI	HYPOTHETICAL	6.49e+00
34	78	12.3	415	1	KPRS_BACSU	RIBOSE PHOSP	6.49e+00
35	78	12.3	600	1	CAIA_ECOLI	PROBABLE CA	6.49e+00
36	78	12.3	450	1	YJE9_YEAST	HYPOTHETICA	6.49e+00
37	78	12.3	466	1	HEMI_MYCTU	GLUTAMYL-TR	6.49e+00
38	77	12.1	138	1	SYT9_MOUSE	SMALL INDUC	8.48e+00
39	77	12.1	154	1	YN9K_YEAST	PUTATIVE MI	8.48e+00
40	77	12.1	180	1	KITH_BPWPV	THYMIDINE K	8.48e+00
41	77	12.1	260	1	ENM_CHLRM	ENDOGLUCANA	8.48e+00
42	77	12.1	444	1	YK34_CAEEL	HYPOTHETICA	8.48e+00
43	77	12.1	612	1	MUS8_STAAU	MODIFICATI	8.48e+00
44	77	12.1	684	1	PCSK_HAEIN	6 PHOSPHOGL	8.48e+00
45	77	12.1	734	1	CPA_MOUSE	21K FINGER	8.48e+00

ALIGNMENTS

RESULT 1
ID YAED_ECOLI STANDARD; PRT; 191 AA.
AC P31546;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHETICAL 21.3 KDA PROTEIN IN ABC-RRSH INTERGEN
GN YAED.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gamma subdivision; ent
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RA Miyamoto K.;
RL Submitted (AUG-1993, to the EMBL/GenBank/DDBJ da
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE; 97426617.
RA Blattner F.R., Plunkett G. III, Bloch C.A., Per
RA Riley M., Collado-Vides J., Glasner F.D., Rode
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia co
RL Science 277:1453-1474(199
RN [3]
RP SEQUENCE FROM N.A.
RA Schramm S., Duncan M., Allen E., Araujo R., Apar
RA Davis K., Federspi
RA Lashkari D., Lew H., Lin S., Namath A., Oefner
RA Davis R.W.;
RL Submitted (SEP-199
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W31
RA Takemoto K., Mori H., Murayama N., Kataoka K.,
RA Yamamoto Y., Inokuchi H., Miki T., Hatada E.,
RA Ichihara S., Mizuno T., Mukai K., Nakata A., Y
RA Mizobuchi K.;
RL Submitted (FEB-199
CC ! SIMILARITY. SIMILAR TO H.INFLUENZAE HI0521.
CC -!- SIMILARITY. TO M.JANNASCHII MJ0015.
CC ! SIMILARITY: TO THE N-TERMINAL PART OF E.COL
CC PHOSPHATE PHOSPHATASE C-MAIN.
CC ----------
CC This SWISS-PROT entry is copyright. It is prod

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; D15061; BAA03661.1; ALT_INIT.
DR   EMBL; AE000129; AAC73311.1; -.
DR   EMBL; U70214; AAB08628.1; -.
DR   EMBL; D83536; BAA77877.1; -.
DR   ECOGENE; EG11736; YAED.
KW   Hypothetical protein.
SQ   SEQUENCE   191 AA;  21294 MW;  E7814B34A23128FA CRC64;

  Query Match           14.7%;  Score 93;  DB 1;  Length 191;
  Best Local Similarity  26.5%;  Pred. No. 3.97e-02;
  Matches   18;  Conservative   21;  Mismatches  24;  Indels   5;  Gaps   5;

Db      68 FETLTEWMDWSLAD-RDVDLDGIYYCPHHPQGSVEEFPQ-VCD-CPKPHPCMLLSARDYL 124
           |  :   |:|  :   :| |:: ||   :: : ::||::  | :  |: | | ::|: |  |
Qy      17 FPAANAWKDCTIQRYEPCPVNNIYTANRNEEITIEEYKVFVNEACK PYP VTLPDRGVL 71

Db     125 HIDMAASY 132
           | :::|
Qy      75 SGDPTSAY 82


RESULT    2
ID   NKC1_SQUAC     STANDARD;      PRT:  1191 AA.
AC   P55013;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   BUMETANIDE-SENSITIVE SODIUM-(POTASSIUM)-CHLORIDE COTRANSPORTER 1
DE   (NA-K-CL SYMPORTER) (NKCC).
GN   SLC12A2 OR NKCC1.
OS   Squalus acanthias (Spiny dogfish).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC   Elasmobranchii; Neoselachii; Squalea; Squaloidei; Squalidae; Squalus.
RN   [1]
RP   SEQUENCE FROM N.A., AND PHOSPHORYLATION SITES.
RC   TISSUE=RECTAL GLAND;
RX   MEDLINE; 94181560.
RA   Xu J.-C., Lytle C., Zhu T.T., Payne J.A., Benz E. Jr., Forbush B. III;
RT   "Molecular cloning and functional expression of the
RT   bumetanide-sensitive Na-K-Cl cotransporter.";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:2201-2205(1994).
CC   -!- FUNCTION: ELECTRICALLY SILENT TRANSPORTER SYSTEM WHICH IS A
CC       MEDIATOR OF SODIUM AND CHLORIDE REABSORPTION. PLAYS A VITAL ROLE
CC       IN THE REGULATION OF IONIC BALANCE AND CELL VOLUME.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN MANY TISSUES.
CC   -!- SIMILARITY: BELONGS TO THE SLC12A FAMILY OF TRANSPORTERS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U05958; AAB60617.1; -.
DR   PFAM; PF00324; aa_permeases; 1.
DR   PRINTS; PR01207; NAKCLIRNSPRT.
DR   PRINTS; PR01208; NAKCLIRSPRT1.
KW   Transport; Transmembrane; Glycoprotein, Phosphorylation.
FT   DOMAIN        1    257       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    258    278       POTENTIAL.
FT   DOMAIN      279    281       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    282    302       POTENTIAL.
FT   DOMAIN      303    339       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    340    360       POTENTIAL.
FT   DOMAIN      361    382       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    383    403       POTENTIAL.
FT   DOMAIN      404    407       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    408    428       POTENTIAL.
FT   DOMAIN      429    458       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    459    479       POTENTIAL.
FT   DOMAIN      480    496       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    497    517       POTENTIAL.
FT   DOMAIN      518    564       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    565    585       POTENTIAL.
FT   DOMAIN      586    630       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    631    651       POTENTIAL.
FT   TRANSMEM    652    672       POTENTIAL.
FT   DOMAIN      673    689       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    690    710       POTENTIAL.
FT   DOMAIN      711    775       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    776    796       POTENTIAL.
FT   DOMAIN      797   1191       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       27     30       POLY-PRO.
FT   DOMAIN       51     57       POLY-GLY.
FT   DOMAIN       67     70       POLY-ALA.
FT   DOMAIN       89     92       POLY-ALA.
FT   DOMAIN       96     99       POLY-ALA.
FT   DOMAIN      126    129       POLY-PRO.
FT   CARBOHYD    523    523       POTENTIAL.
FT   CARBOHYD    535    535       POTENTIAL.
FT   CARBOHYD    546    546       POTENTIAL.
FT   MOD_RES     189    189       PHOSPHORYLATION.
FT   MOD_RES    1114   1114       PHOSPHORYLATION.
SQ   SEQUENCE   1191 AA;  129774 MW;  BA7BB98154315000 CRC64;

  Query Match           14.7%;  Score 93;  DB 1;  Length 1191;
  Best Local Similarity  25.5%;  Pred. No. 3.97e-02;
  Matches   13;  Conservative   14;  Mismatches  23;  Indels   2;  Gaps   2;

Db    1014 LTLLIPYILTTKKKWKDCKIRVFIGGKINPIHHDERTMATLLSKFRIDFSD 1064
           |  | :|: :::  |||| |: :  | :| |     |      :  ::: | :
Qy      36 FIALAAYFTAANA-WKDCLIGKKMGGVNNLYIANENEEIIIEEYKV-FVN 58


RESULT    3
ID   NKC1_MOUSE     STANDARD;      PRT:  1205 AA.
AC   P55012;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   BUMETANIDE-SENSITIVE SODIUM-(POTASSIUM)-CHLORIDE COTRANSPORTER 1
DE   (BASOLATERAL NA-K-CL SYMPORTER).
GN   SLC12A2 OR NKCC1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95014374.
RA   Delpire E., Rauchman M.I., Beier D.R., Hebert S.C., Gullans S.R.;
RT   "Molecular cloning and chromosome localization of a putative
RT   basolateral Na(+)-K(+)-2Cl cotransporter from mouse inner medullary
RT   collecting duct (mIMCD-3) cells.";
RL   J. Biol. Chem. 269:25677-25683(1994).
CC   -!- FUNCTION: ELECTRICALLY SILENT TRANSPORTER SYSTEM WHICH IS A
CC       MEDIATOR OF SODIUM AND CHLORIDE REABSORPTION. PLAYS A VITAL ROLE
CC       IN THE REGULATION OF IONIC BALANCE AND CELL VOLUME.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE SLC12A FAMILY OF TRANSPORTERS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U18174; AAC77832.1; -.
DR   MGD; MGI:101924; SLC12A2.
DR   PFAM; PF00324; aa_permeases; 1.
DR   PRINTS; PR01207; NAKCLTRNSPRT.
DR   PRINTS; PR01208; NAKCLTRSPRTI.
KW   Transport; Transmembrane; Glycoprotein.
FT   DOMAIN        1    278       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    279    299       POTENTIAL.
FT   DOMAIN      300    302       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    303    323       POTENTIAL.
FT   DOMAIN      324    360       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    361    381       POTENTIAL.
FT   DOMAIN      382    403       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    404    424       POTENTIAL.
FT   DOMAIN      425    428       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    429    449       POTENTIAL.
FT   DOMAIN      450    479       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    480    500       POTENTIAL.
FT   DOMAIN      501    513       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    514    534       POTENTIAL.
FT   DOMAIN      535    584       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    585    605       POTENTIAL.
FT   DOMAIN      606    650       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    651    671       POTENTIAL.
FT   TRANSMEM    672    692       POTENTIAL.
FT   DOMAIN      693    709       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    710    730       POTENTIAL.
FT   DOMAIN      731    892       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    893    913       POTENTIAL.
FT   DOMAIN      914   1205       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       90    107       POLY-ALA.
FT   CARBOHYD    546    546       POTENTIAL.
FT   CARBOHYD    555    555       POTENTIAL.
SQ   SEQUENCE   1205 AA;  130949 MW;  A93D2A0468377120 CRC64;

  Query Match           14.7%;  Score 93;  DB 1;  Length 1205;
  Best Local Similarity  25.5%;  Pred. No. 3.97e-02;
  Matches   13;  Conservative   14;  Mismatches  23;  Indels   2;  Gaps   2;

Db   1028 LTLLIPYLLTTKKKWKDCKIPVFIGGKINRIDHDRRAMATLLSKPRIDFSD 1078
          |  | :|: ::: |||| |: : | :| |    |     : ::: | :
Qy     10 LIALAAYFPAANA-WKDCIIQRYKDGDVNNIYTANRNEEITIEEYKV-FVN 58


RESULT    4
ID   NKC1_HUMAN     STANDARD;      PRT;  1212 AA.
AC   P55011;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   HUMETANIDE-SENSITIVE SODIUM-(POTASSIUM)-CHLORIDE COTRANSPORTER 1
DE   (BASOLATERAL NA-K-CL SYMPORTER).
GN   SLC12A2 OR NKCC1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=COLON;
RX   MEDLINE; 95355397.
RA   Payne J.A., Xu J.-C., Haas M., Lytle Y.C., Ward D., Forbush B. III;
RT   "Primary structure, functional expression, and chromosomal
RT   localization of the bumetanide-sensitive Na-K-Cl cotransporter in
RT   human colon.";
RL   J. Biol. Chem. 270:17977-17985(1995).
CC   -!- FUNCTION: ELECTRICALLY SILENT TRANSPORTER SYSTEM WHICH IS A
CC       MEDIATOR OF SODIUM AND CHLORIDE REABSORPTION. PLAYS A VITAL ROLE
CC       IN THE REGULATION OF IONIC BALANCE AND CELL VOLUME.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN MANY TISSUES.
CC   -!- SIMILARITY: BELONGS TO THE SLC12A FAMILY OF TRANSPORTERS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced
CC   between the Swiss Institute of Bioinformatics an
CC   the European Bioinformatics Institute. There are
CC   use by non-profit institutions as long as its
CC   modified and this statement is not removed. Usage
CC   entities requires a license agreement (See http://
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U30246; AAC50561.1; -.
DR   MIM; 600840; -.
DR   PFAM; PF00324; aa_permeases; 1.
DR   PRINTS; PR01207; NAKCLTRNSPRT.
DR   PRINTS; PR01208; NAKCLTRSPRTI.
KW   Transport; Transmembrane; Glycoprotein.
FT   DOMAIN        1     92       CYTOPLASMIC (POTENTIA
FT   TRANSMEM     93    113       POTENTIAL.
FT   DOMAIN      114    285       EXTRACELLULAR (POTENT
FT   TRANSMEM    286    306       POTENTIAL.
FT   DOMAIN      307    309       CYTOPLASMIC (POTENTIA
FT   TRANSMEM    310    330       POTENTIAL.
FT   DOMAIN      331    366       EXTRACELLULAR (POTENT
FT   TRANSMEM    367    387       POTENTIAL.
FT   DOMAIN      388    413       CYTOPLASMIC (POTENTIA
FT   TRANSMEM    411    431       POTENTIAL.
FT   DOMAIN      432    435       EXTRACELLULAR (POTENT
FT   TRANSMEM    436    456       POTENTIAL.
FT   DOMAIN      457    486       CYTOPLASMIC (POTENTIA
FT   TRANSMEM    487    507       POTENTIAL.
FT   DOMAIN      508    520       EXTRACELLULAR (POTENT
FT   TRANSMEM    521    541       POTENTIAL.
FT   DOMAIN      542    591       CYTOPLASMIC (POTENTIA
FT   TRANSMEM    592    612       POTENTIAL.
FT   DOMAIN      613    657       EXTRACELLULAR (POTENT
FT   TRANSMEM    658    678       POTENTIAL.
FT   TRANSMEM    679    699       POTENTIAL.
FT   DOMAIN      700    724       EXTRACELLULAR (POTENT
FT   TRANSMEM    725    745       POTENTIAL.
FT   DOMAIN      746   1212       CYTOPLASMIC (POTENTIA
FT   DOMAIN       93    111       POLY-ALA.
FT   DOMAIN      188    194       POLY-GLY.
FT   CARBOHYD    168    168       POTENTIAL.
SQ   SEQUENCE   1212 AA;  [illegible] MW;  A6BF174ACCBA2EC1

  Query Match           14.7%;  Score 93;  DB 1;  Len
  Best Local Similarity  25.5%;  Pred. No. 3.97e-02;
  Matches   13;  Conservative   14;  Mismatches  23;

Db   1035 LTLLIPYLLTTKKKWK [illegible] KINRIDHDRRAMATLLSK
          |  | :|: ::: |||| |: : | :| |    |     : ::: | :
Qy     10 LIALAAYFPAANA-WK [illegible] VNNIYTANRNEEITIEEYKV-FVN 58


RESULT    5
ID   NTPI_FOWPV     STANDARD;      PRT;   637 AA.
AC   O72907;
DT   15-FEB-2000 (Rel. 39, Created)
DT   15-FEB-2000 (Rel. 39, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   NUCLEOSIDE TRIPHOSPHATASE I (E.C.3.6.1.15) (NTPase
DE   PHOSPHOHYDROLASE I) (NPH I).
GN   FP-D11.
OS   Fowlpox virus.
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae;
OC   Avipoxvirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN FP9 / HP-44;
RA   Skinner M.A.;
RL   Submitted (MAY 1998) to the EMBL/GenBank/DDBJ dat
CC   -!- FUNCTION: SERVES TWO ROLES IN TRANSCRIPTION:
CC       WITH VIRAL TERMINATION FACTOR/CAPPING ENZYME
CC       OF ELONGING CONTAINING NASCENT RNA FROM THE E
```

```
CC       IT ACTS BY ITSELF AS A POLYMERASE ELONGATION FACTOR TO FACILITATE
CC       READTHROUGH OF INTRINSIC PAUSE SITES (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: NTP + H(2)O = NDP + ORTHOPHOSPHATE.
CC   -!- SIMILARITY: BELONGS TO THE NPH I SUBFAMILY OF HELICASES
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ---------- ---------- ----------
DR   EMBL; AJ005163; CAA06401.1; -.
DR   PFAM; PF00176; SNF2_N; 1.
DR   PFAM; PF00271; helicase_C; 1.
KW   Hydrolase; ATP binding; Transcription; Late protein.
FT   DOMAIN        35    275       SNF2_N.
FT   DOMAIN       397    483       HELICASE_C.
FT   NP_BIND       56     63       ATP (BY SIMILARITY).
FT   SITE         142    145       DEXH BOX.
SQ   SEQUENCE   637 AA;  73774 MW;  9FB38D2B0519F9D8 CRC64;

  Query Match           14.5%;  Score 92;  DB 1;  Length 637;
  Best Local Similarity  31.0%;  Pred. No. 5.73e-02;
  Matches   13;  Conservative   11;  Mismatches  15;  Indels   3;  Gaps   3;

Db     108 KDCTIMNYDLQNFHNKFFNIKS-NVKS-RIFLTDVCHNF 147
           ||||| | | : :| : :| : | :  ::|:  : || :
Qy      24 KDCIIQRYKDGDVNNIYTANRNEEITIEEYKVFVN-EACHPY 64


RESULT    6
ID   CG23_YEAST     STANDARD;      PRT;   427 AA.
AC   P24870;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   G2/MITOTIC-SPECIFIC CYCLIN 3.
GN   CLB3 OR YDL155W OR D1539.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C;
RX   MEDLINE; 93051823.
RA   Richardson H., Lew D.J., Henze M., Sugimoto K., Reed S.I.;
RT   "Cyclin-B homologs in Saccharomyces cerevisiae function in S phase
RT   and in G2.";
RL   Genes Dev. 6:2021-2034(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92385891.
RA   Fitch I., Dahmann C., Surana U., Amon A., Nasmyth K., Goetsch L.,
RA   Byers B., Futcher B.;
RT   "Characterization of four B-type cyclin genes of the budding yeast
RT   Saccharomyces cerevisiae.";
RL   Mol. Biol. Cell 3:805-818(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN S288C / DY1670.
RX   MEDLINE; 97127830.
RA   Dolaveau T., Bluqeon C., Jacq C., Perea J.;
RT   "Analysis of a 23 kb region on the left arm of yeast chromosome IV.";
RL   Yeast 12:1587-1592(1996).
RN   [4]
RP   SEQUENCE OF 194-383 FROM N.A.
RX   MEDLINE; 91191554.
RA   Surana U., Robitsch H., Price C., Schuster T., Fitch I., Futcher A.B.,
RA   Nasmyth K.;
RT   "The role of CDC28 and cyclins during mitosis in the budding yeast S.
RT   cerevisiae.";
RL   Cell 65:145-161(1991).
CC   -!- FUNCTION: ESSENTIAL FOR THE CONTROL OF THE CELL CYCLE AT THE G2/M
CC       (MITOSIS) TRANSITION. INTERACTS WITH THE CDC2 PROTEIN KINASE TO
CC       FORM MPF. G2/M CYCLINS ACCUMULATE STEADILY DURING G2 AND ARE
CC       ABRUPTLY DESTROYED AT MITOSIS.
CC   -!- SIMILARITY: BELONGS TO THE CYCLIN FAMILY. CYCLIN AB SUBFAMILY.
CC   ---------- ---------- ---------- ----------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ---------- ---------- ----------
DR   EMBL; X69425; CAA49201.1; -.
DR   EMBL; M80302; AAA34765.1; -.
DR   EMBL; X97751; CAA66336.1; -.
DR   EMBL; Z74203; CAA98729.1; -.
DR   PIR; S14167; S14167.
DR   PIR; A46228; A46228.
DR   PIR; A60048; A60048.
DR   HSSP; P30274; 1VIN.
DR   SGD; L0000351; CLB3.
DR   PFAM; PF00134; cyclin; 1.
DR   PROSITE; PS00292; CYCLINS; 1.
KW   Cyclin; Cell cycle; Cell division; Mitosis; Multigene family.
FT   CONFLICT     48     48       R -> D (IN REF. 2 AND CAA98729).
SQ   SEQUENCE   427 AA;  49516 MW;  2DC484475876672B CRC64;

  Query Match           14.0%;  Score 89;  DB 1;  Length 427;
  Best Local Similarity  24.6%;  Pred. No. 1.70e-01;
  Matches   15;  Conservative   22;  Mismatches  17;  Indels   7;  Gaps   5;

Db     235 KEVVFVNKFQLVCAASLFLAAKYEEINCPLKDFVYMSENCYSKNDLLDAERTILNGLE 293
           |: : : ::       ::: | : |||   || :: |:::| | :   :| ::|::|:| :
Qy      24 KDCIIQRYKDGDVNNIYTANRNEEI  TIEEYKVFVNEACHPYPVIL-PDPSVLSG-D 77

Db     294 F 294
           |
Qy      78 F 78


RESULT    7
ID   BLAI_STAAU     STANDARD;      PRT;   126 AA.
AC   P18415;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   PENICILLINASE REPRESSOR (REGULATORY PROTEIN BLAI) (BETA-LACTAMASE
DE   REPRESSOR PROTEIN).
GN   BLAI OR PENI.
OS   Staphylococcus aureus.
OG   Plasmid pI258.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Staphylococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 9789; TRANSPOSON=TN552;
RX   MEDLINE; 91014696.
RA   Rowland S.J., Dyke K.G.H.;
RT   "Tn552, a novel transposable element from Staphylococcus aureus.";
RL   Mol. Microbiol. 4:961-975(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   PLASMID=PI258;
RX   MEDLINE; 91319567.
RA   Wang P.Z., Projan S.J., Novick R.P.;
RT   "Nucleotide sequence of beta lactamase regulatory genes from
RT   staphylococcal plasmid pI258.";
RL   Nucleic Acids Res. 19:4000-4000(1991).
RN   [3]
RP   SEQUENCE FROM N.A.
```

```
RX   MEDLINE; 93073773.
RA   Smith M.C., Murray B.E.;
RT   "Sequence analysis of the beta-lactamase repressor from
RT   Staphylococcus aureus and hybridization studies with two
RT   beta-lactamase-producing isolates of Enterococcus faecalis.";
RL   Antimicrob. Agents Chemother. 36:2265-2269(1992).
CC   -!- FUNCTION: REPRESSION OF BETA-LACTAMASE.
CC   -!- SIMILARITY: BELONGS TO THE BLAI/MECI FAMILY OF TRANSCRIPTIONAL
CC       REGULATORS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X52734; CAA36951.1; -.
DR   EMBL; M62650; AAA26603.1; -.
DR   EMBL; M92376; AAA26605.1; -.
DR   PIR; S34446; S34446.
DR   PIR; S11782; S11782.
KW   Transposable element; Transcription regulation; Repressor;
KW   DNA-binding; Plasmid.
SQ   SEQUENCE   126 AA;  14876 MW;  E0759E83A5183117 CRC64;

  Query Match           13.9%;  Score 88;  DB 1;  Length 126;
  Best Local Similarity  33.3%;  Pred. No. 2.42e-01;
  Matches   12;  Conservative   12;  Mismatches   9;  Indels   3,  Gaps   3,

Db      55 KE-TIKRYASENIY-EYSSNIKEDDIKMKIAKTFLN 88
           |: || |||  ::  :|:|| ::::| :   | |:|
Qy      24 KDCIIQRYKDGEVNNIYTAN ENEEITIEEYKVFVN 58


RESULT    8
ID   NIFN_SYNP8      STANDARD;      PRT;   461 AA.
AC   Q07456;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   NITROGENASE IRON-MOLYBDENUM COFACTOR BIOSYNTHESIS PROTEIN NIFN.
GN   NIFN.
OS   Synechococcus sp. (strain PCC 6301 / SF-1) (Cyanothece PCC 8801).
OC   Bacteria; Cyanobacteria; Chroococcales; Cyanobacteria
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Chen H.M.;
RT   "NifENXW operon.";
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THIS PROTEIN MAY PLAY A ROLE IN THE BIOSYNTHESIS OF
CC       THE PROSTHETIC GROUP OF NITROGENASE (FEMO COFACTOR).
CC   -!- PATHWAY: FE-MO COFACTOR BIOSYNTHESIS.
CC   -!- SIMILARITY: BELONGS TO THE NIFE/NIFK/NIFN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF003700; AAC35195.1; -.
DR   PFAM; PF00148; oxidored_nitro; 1.
DR   PROSITE; PS00699; NITROGENASE_1_1; 1.
DR   PROSITE; PS00090; NITROGENASE_1_2; FALSE_NEG.
KW   Nitrogen fixation.
SQ   SEQUENCE   461 AA;  50377 MW;  047765AA3BBB0A2F CRC64;

  Query Match           13.6%;  Score 86;  DB 1;  Length 461;
  Best Local Similarity  33.3%;  Pred. No. 4.89e-01;
  Matches   16;  Conservative   14;  Mismatches   15;

Db     138 YK-GSLQDCYAAT-VEQIVATLYNAFIAENARS-AVIYP-P-VIV
           || | :: |:|   :|:  ::: |: : :: :|| |: : :
Qy      41 YKDGDVNNIYTANRNEEITIEEYFVFVNEACHPYPVILPL-E-SV


RESULT    9
ID   VGF1_IBVB       STANDARD;      PRT:  3951 AA.
AC   P27920;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   01-AUG-1992 (Rel. 23, Last annotation update)
DE   F1 PROTEIN.
GN   F1.
OS   Avian infectious bronchitis virus (strain Beaudette
OC   Viruses; ssRNA positive-strand viruses, no DNA sta
OC   Coronaviridae; Coronavirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 87111468.
RA   Boursnell M.E.G., Brown T.D.K., Foulds I.J., Green
RA   Binns M.M.;
RT   "Completion of the sequence of the genome of the
RT   infectious bronchitis virus.";
RL   J. Gen. Virol. 68:57-77(1987).
CC   ---------- ------- -- - ------ -
CC   This SWISS-PROT entry is copyright. It is produced
CC   between the Swiss Institute of Bioinformatics an
CC   the European Bioinformatics Institute. There are
CC   use by non-profit institutions as long as its
CC   modified and this statement is not removed. Usage
CC   entities requires a license agreement (See http://
CC   or send an email to license@isb-sib.ch).
CC   ---------- ---------- ------ - ------- -
DR   EMBL; M94356; AAA46225.1; -.
DR   EMBL; M95169; AAA76243.1; -.
DR   PIR; A33094; VFIHB1.
DR   PFAM; PF01661; DUF27; 1.
SQ   SEQUENCE   3951 AA;  44112 MW;  9B8B2E7E2545F

  Query Match           13.6%;  Score 86;  DB 1;  Length
  Best Local Similarity  31.0%;  Pred. No. 4.89e-01;
  Matches    9;  Conservative   7;  Mismatches   13;

DB    2742 YLQACRAWLAYALLQYENSDVEIVVTIIR 2770
           |: | ||   :::|::: :: :|| |
Qy      16 YEIAANAWKDCIIQRYEDCEVNNLYIANR 44


RESULT   10
ID   COX3_STRPU      STANDARD;      PRT:   260 AA.
AC   P15546;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   CYTOCHROME C OXIDASE POLYPEPTIDE III (EC 1.9.3.1)
GN   COIII.
OS   Strongylocentrotus purpuratus (Purple sea urchin)
OC   Mitochondrion.
OC   Eukaryota; Metazoa; Echinodermata; Echinozoa; Ech
OC   Echinoidea; Euechinoidea; Echinacea; Strongylocent
OC   Strongylocentrotus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 89011951.
RA   Jacobs H.T., Elliott D.J., Math V.B., Farquharson
RT   "Nucleotide sequence and gene organization of sea
RT   mitochondrial DNA.";
RL   J. Mol. Biol. 202:185-217(1988).
CC   -!- FUNCTION: SUBUNIT 1, II, AND III FORM THE FUNC
CC       THE ENZYME COMPLEX.
CC   -!- CATALYTIC ACTIVITY: 4 FERRO*CYTOCHROME C + O(2)
```

```
CC       4 FERRICYTOCHROME C.
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 3 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X12631; -; NOT_ANNOTATED_CDS.
DR   PIR; S01506; S01506.
DR   HSSP; P00415; 1OCC.
DR   PFAM; PF00510; COX3; 1.
KW   Oxidoreductase; Mitochondrion; Transmembrane.
SQ   SEQUENCE   260 AA;  29489 MW;  74166704 2F81A00 CRC64;

  Query Match           13.4%;  Score 85;  DB 1;  Length 260;
  Best Local Similarity  39.4%;  Pred. No. 6.93e-01;
  Matches   13;  Conservative    7;  Mismatches  11;  Indels   2;  Gaps   2;

Db    159 IQALFLF-VALCSYFTALQAW-EYIDAPFTIAD 189
          :|| | | :||::|| |  || : |   :  :|
Qy      3 VQAIFATLIALAAYFPAANAWKDCIIQRYKDGD 35


RESULT   11
ID   CPT1_HUMAN     STANDARD;      PRT;   773 AA.
AC   P50416;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   CARNITINE O-PALMITOYLTRANSFERASE I, MITOCHONDRIAL LIVER ISOFORM
DE   (EC 2.3.1.21) (CPT I) (CPTI L).
GN   CPT1A OR CPT1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 95199277.
RA   Britton C.H., Schultz R.A., Zhang B., Esser V., Foster D.W.,
RA   McGarry J.D.;
RT   "Human liver mitochondrial carnitine palmitoyltransferase I:
RT   characterization of its cDNA and chromosomal localization and partial
RT   analysis of the gene.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:1984-1988(1995).
CC   -!- CATALYTIC ACTIVITY: PALMITOYL-COA + L-CARNITINE   COA +
CC       L-PALMITOYLCARNITINE.
CC   -!- ENZYME REGULATION: INHIBITORS SUCH AS MALONYL-COA INTERACT WITH
CC       ITS CATALYTIC DOMAIN AND NOT WITH AN ASSOCIATED REGULATORY
CC       COMPONENT.
CC   -!- PATHWAY: INVOLVED IN FATTY ACID BETA-OXIDATION.
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL OUTER MEMBRANE.
CC   -!- TISSUE SPECIFICITY: STRONG EXPRESSION IN KIDNEY AND HEART, AND
CC       LOWER IN LIVER AND SKELETAL MUSCLE.
CC   -!- SIMILARITY: BELONGS TO THE CARNITINE/CHOLINE ACETYLTRANSFERASE
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L39211; AAC41748.1; -.
DR   MIM; 600528; -.
DR   MIM; 255120; -.
DR   PFAM; PF00755; Carn_acyltransf; 1.
DR   PROSITE; PS00439; ACYLTRANSF_C_1; 1.
DR   PROSITE; PS00440; ACYLTRANSF_C_2; 1.
KW   Transferase; Acyltransferase; Mitochondrion; Outer membrane;
KW   Fatty acid metabolism; Transport; Transmembrane; Multigene family.
FT   DOMAIN       1     47       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    48     73       POTENTIAL.
FT   DOMAIN      74    102       MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
FT   TRANSMEM   103    122       POTENTIAL.
FT   DOMAIN     123    773       CYTOPLASMIC (POTENTIAL).
FT   ACT_SITE   473    473       POTENTIAL.
SQ   SEQUENCE   773 AA;  88428 MW;  44639F228343C990 CRC64;

  Query Match           13.4%;  Score 85;  DB 1;  Length 773;
  Best Local Similarity  32.1%;  Pred. No. 6.93e-01;
  Matches    9;  Conservative   11;  Mismatches   7;  Indels   1;  Gaps   1;

Db     32 YLSGLHSWKKKF1-RFKNGTTGVYPAS 58
          |::: ::|| :| |:|:| :  :| |:
Qy     16 YFPAANAWKDCIIQRYKDGDVNNIYTAN 43


RESULT   12
ID   PIP5_HUMAN     STANDARD;      PRT;  1252 AA.
AC   P16885;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   1-PHOSPHATIDYLINOSITOL-4,5-BISPHOSPHATE PHOSPHODIESTERASE GAMMA 2
DE   (EC 3.1.4.11) (PLC-GAMMA-2) (PHOSPHOLIPASE C-GAMMA-2) (PLC-IV).
GN   PLCG2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LYMPHOBLAST;
RX   MEDLINE; 89078616.
RA   Ohta S., Matsui A., Nazawa Y., Kaqawa Y.;
RT   "Complete cDNA encoding a putative phospholipase C from transformed
RT   human lymphocytes.";
RL   FEBS Lett. 242:31-35(1988).
CC   -!- FUNCTION: THE PRODUCTION OF THE SECOND MESSENGER MOLECULES
CC       DIACYLGLYCEROL (DAG) AND INOSITOL 1,4,5-TRISPHOSPHATE (IP3) IS
CC       MEDIATED BY ACTIVATED PHOSPHATIDYLINOSITOL-SPECIFIC PHOSPHOLIPASE
CC       C ENZYMES. IT IS A CRUCIAL ENZYME IN TRANSMEMBRANE SIGNALING.
CC   -!- CATALYTIC ACTIVITY: 1-PHOSPHATIDYL-D-MYO-INOSITOL 4,5-BISPHOSPHATE
CC       + H(2)O   D-MYO-INOSITOL 1,4,5 TRISPHOSPHATE + DIACYLGLYCEROL.
CC   -!- COFACTOR: REQUIRES CALCIUM.
CC   -!- PTM: THE RECEPTOR-MEDIATED ACTIVATION OF PLC-GAMMA 1 AND 2
CC       INVOLVES THEIR PHOSPHORYLATION BY TYROSINE KINASES IN RESPONSE
CC       TO LIGATION OF A VARIETY OF GROWTH FACTOR RECEPTORS AND IMMUNE
CC       SYSTEM RECEPTORS.
CC   -!- MISCELLANEOUS: THERE ARE AT LEAST SIX FORMS OF PLC ENZYMES.
CC   -!- SIMILARITY: DOMAINS X AND Y ARE CONSERVED IN DIFFERENT FORMS
CC       OF PLC AND ARE ESSENTIAL FOR CATALYTIC ACTIVITY.
CC   -!- SIMILARITY: CONTAINS 1 C2 DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 SH2 DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X14034; CAA32194.1; -.
DR   EMBL; M37238; AAA60112.1; -.
DR   PIR; S02004; S02004.
DR   HSSP; P08487; 2PLE.
DR   MIM; 600220; -.
```

```
DR   PFAM; PF00168; C2; 1.
DR   PFAM; PF00169; PH; 3.
DR   PFAM; PF00388; PI-PLC-X; 1.
DR   PFAM; PF00387; PI-PLC-Y; 1.
DR   PFAM; PF00017; SH2; 2.
DR   PFAM; PF00018; SH3; 1.
DR   PRINTS; PR00390; PHPHLIPASEC.
DR   PRINTS; PR00401; SH2DOMAIN.
DR   PRINTS; PR00452; SH3DOMAIN.
DR   PROSITE; PS50001; SH2; 2.
DR   PROSITE; PS50002; SH3; 1.
DR   PROSITE; PS50003; PH_DOMAIN; 1.
DR   PROSITE; PS50004; C2_DOMAIN_2; 1.
DR   PROSITE; PS50007; PIPLC_X_DOMAIN; 1.
DR   PROSITE; PS50008; PIPLC_Y_DOMAIN; 1.
KW   Hydrolase; Lipid degradation; Transducer; SH2 domain; SH3 domain;
KW   Calcium binding; Phosphorylation.
FT   DOMAIN       20    131       PH.
FT   DOMAIN      312    456       DOMAIN X.
FT   DOMAIN      542    635       SH2.
FT   DOMAIN      646    735       SH2.
FT   DOMAIN      769    829       SH3.
FT   DOMAIN      930   1044       DOMAIN Y.
FT   DOMAIN     1059   1152       C2 DOMAIN.
FT   ACT_SITE    327    327       BY SIMILARITY.
FT   ACT_SITE    372    372       BY SIMILARITY.
SQ   SEQUENCE   1252 AA;  146119 MW;  6F63E35162DFB664 CRC64;

  Query Match           13.2%;  Score 84;  DB 1;  Length 1252;
  Best Local Similarity  19.1%;  Pred. No. 9.77e-01;
  Matches   13;  Conservative   23;  Mismatches  28,  Indels   4,  Gaps   4,

Db    710 FESLVELVSYYEKHSLYRK-MRLRYPVTPELLERYNTERDINSLYDVSRMYV-DPSEINP 767
          | ::|: | :|:   :  :  :  ||    :: : | ::|: :  :  :::| ::    |
Qy      7 FATLIALAAYFPAANAWKDCIIQRYK-DGDVNNIYIANRNEEIIIEEYKVEVNEACHPYP 65

Db    768 S-MPQRTV 774
           :|:|:|
Qy     66 VIIPDRSV 73


RESULT   13
ID   SCPA_PENSP     STANDARD;      PRT:   192 AA.
AC   P02636;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   SARCOPLASMIC CALCIUM-BINDING PROTEIN, ALPHA-B AND -A CHAINS (SCP ALPHA
DE   CHAIN).
OS   Penaeus sp. (Penoeid shrimp).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Dendrobranchiata; Penaeoidea;
OC   Penaeidae; Penaeus.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 84289338.
RA   Takagi T., Konishi K.;
RT   "Amino acid sequence of alpha chain of sarcoplasmic calcium binding
RT   protein obtained from shrimp tail muscle.";
RL   J. Biochem. 95:1603-1615(1984).
CC   -!- FUNCTION: LIKE PARVALBUMINS, SCP'S SEEM TO BE MORE ABUNDANT IN
CC       FAST CONTRACTING MUSCLES, BUT NO FUNCTIONAL RELATIONSHIP CAN BE
CC       ESTABLISHED FROM THIS DISTRIBUTION.
CC   -!- SUBUNIT: SCPS FROM CRAYFISH, LOBSTER, AND SHRIMP ARE POLYMORPHIC
CC       DIMERS; THREE ISOTYPES (ALPHA-ALPHA, ALPHA-BETA, AND BETA-BETA)
CC       HAVE BEEN IDENTIFIED.
CC   -!- MISCELLANEOUS: THE SARCOPLASMIC CALCIUM-BINDING PROTEINS ARE
CC       ABUNDANT IN THE MUSCLE OF ARTHROPODS, MOLLUSCS, ANNELIDS, AND
CC       PROTOCHORDATES.
CC   -!- MISCELLANEOUS: THIS PROTEIN HAS THREE FUNCTIONAL CALCIUM-BINDING
CC       SITES; POTENTIAL SITE 4 HAS LOST AFFINITY FOR CALCIUM.
CC   -!- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS
DR   PIR; A03073; KLSSAB.
DR   PIR; A03074; KLSSAA.
DR   PFAM; PF00036; efhand; 2.
DR   PROSITE; PS00018; EF_HAND; 4.
KW   Muscle protein; Calcium-binding; Duplication; Acet
FT   MOD_RES       1       1       ACETYLATION.
FT   CA_BIND      17      28       SITE 1.
FT   CA_BIND      69      80       SITE 2.
FT   CA_BIND     113     124       SITE 3.
FT   DOMAIN      140     151       ANCESTRAL CALCIUM SI
FT   VARIANT       4       4       MISSING (IN ALPHA-A).
FT   VARIANT      61      61       MISSING (IN ALPHA-A).
FT   VARIANT      95      95       S -> G (IN ALPHA-A).
FT   VARIANT     107     107       N -> D (IN ALPHA-A).
SQ   SEQUENCE   192 AA;  21979 MW;  B7FDCBBEB79A2727 CR

  Query Match           12.9%;  Score 82;  DB 1;  Leng
  Best Local Similarity  27.9%;  Pred. No. 1.93e+00;
  Matches   15;  Conservative   17;  Mismatches  22;

Db     97 TLIESPGELPSAADYANNKFIMDNLNNELAELADFNKDGEVTVDEI
          |||   : |:||:  ::  | |    :: ::   |:: |:|::|
Qy      9 TLIALAAYFPAANAWKDCIIQRYKDGDVNNIYIANRNEEIIIEEY

RESULT   14
ID   CAO2_CANMA     STANDARD;      PRT:   724 AA.
AC   Q00468;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   ACYL-COENZYME A OXIDASE 2 (EC 1.3.3.6) (ACYL-CO
GN   POX2.
OS   Candida maltosa (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Sac
OC   anamorphic Saccharomycetales; Candida.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IAM 12247;
RX   MEDLINE; 96144267.
RA   Masuda Y., Park S.M., Ohta A., Takagi M.;
RT   "Cloning and characterization of the POX2 gene in
RL   Gene 167:157-161(1995).
CC   -!- CATALYTIC ACTIVITY: ACYL-S-A + O(2) = TRANS-2,
CC       H(2)O(2) (ACTION ON SOA DERIVATIVES OF FATTY AC
CC       FROM 8 TO 18).
CC   -!- COFACTOR: FAD.
CC   -!- PATHWAY: INITIAL STEP OF THE PEROXISOMAL FATTY
CC       SYSTEM.
CC   -!- SUBUNIT: HOMOOCTAMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: PEROXISOMAL.
CC   ---------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced
CC   between the Swiss Institute of Bioinformatics an
CC   the European Bioinformatics Institute. There are
CC   use by non-profit institutions as long as its
CC   modified and this statement is not removed. Usage
CC   entities requires a license agreement (See http://
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------
DR   EMBL; D21228; BAA04761.1; -.
DR   PFAM; PF01756; ACOX; 1.
KW   Oxidoreductase; Fatty acid metabolism; Peroxisome;
KW   Multigene family.
SQ   SEQUENCE   724 AA;  [illegible] MW;  [illegible] CRC64;

  Query Match           12.9%;  Score 82;  DB 1;  Leng
  Best Local Similarity  28.9%;  Pred. No. 1.93e+00;
  Matches   13;  Conservative   11;  Mismatches  20;

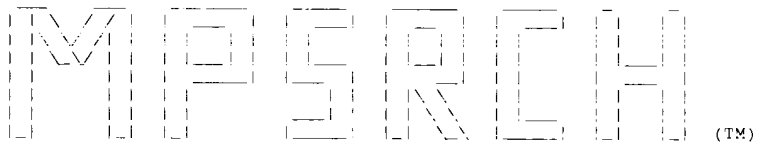
Db    647 VIGLTDSPQQFDNFINSLIGKY DENVYTNYLTNVTNVNDFTNYK
          :|:|:  |  ::   : :| || ||:|  | :|  :    :||
Qy     10 LIALAAYFPAANAWKDCIIQRYKDGDVNNIYTANRNEEITIEEYK
```

```
RESULT   15
ID   CARB_YEAST     STANDARD;      PRT;  1118 AA.
AC   P03965;
DT   23-OCT-1986 (Rel. 02, Created)
DT   23-OCT-1986 (Rel. 02, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   CARBAMOYL-PHOSPHATE SYNTHASE, ARGININE-SPECIFIC, LARGE CHAIN
DE   (EC 6.3.5.5) (ARGININE-SPECIFIC CARBAMOYL-PHOSPHATE SYNTHETASE,
DE   AMMONIA CHAIN).
GN   CPA2 OR YJR109C OR J2002.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 84061899.
RA   Lusty C.J., Widgren E.E., Broglie K.E., Nyunoya H.;
RT   "Yeast carbamyl phosphate synthetase. Structure of the yeast gene and
RT   homology to Escherichia coli carbamyl phosphate synthetase.";
RL   J. Biol. Chem. 258:14466-14477(1983).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Ramezani Rad M., Kirchrath L., Hollenberg C.P.;
RL   Submitted (SEP 1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 1-859 FROM N.A.
RA   Rose M., Koetter P., Entian K.D.;
RL   Submitted (SEP 1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: 2 ATP + GLUTAMINE + CO(2) + H(2)O = 2 ADP +
CC       ORTHOPHOSPHATE + GLUTAMATE + CARBAMOYL PHOSPHATE.
CC   -!- PATHWAY: BIOSYNTHESIS OF ARGININE.
CC   -!- SUBUNIT: COMPOSED OF TWO CHAINS; THE SMALL (OR GLUTAMINE) CHAIN
CC       PROMOTES THE HYDROLYSIS OF GLUTAMINE TO AMMONIA, WHICH IS USED
CC       BY THE LARGE (OR AMMONIA) CHAIN TO SYNTHESIZE CARBAMOYL PHOSPHATE.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- MISCELLANEOUS: IN EUKARYOTES THIS ENZYME IS SYNTHESIZED BY TWO
CC       PATHWAY-SPECIFIC (ARGININE AND PYRIMIDINE) UNDER SEPARATE CONTROL.
CC   -!- SIMILARITY: HOMOLOGOUS TO THE PROKARYOTE ENZYME.
CC   ---------- ---------- ---------- ---------- ----------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ---------- ---------- ---------- ---------- ----------
DR   EMBL; K01178; AAA66902.1; -.
DR   EMBL; Z49609; CAA89639.1; -.
DR   PIR; A01199; SYBYCP.
DR   HSSP; P00968; 1JDB.
DR   SGD; L0000400; CPA2.
DR   PFAM; PF00289; CPSase_L_chain; 2.
DR   PRINTS; PR00098; CPSASE.
DR   PROSITE; PS00866; CPSASE_1; 2.
DR   PROSITE; PS00867; CPSASE_2; 2.
KW   Arginine biosynthesis; Ligase; Duplication; ATP binding.
FT   REPEAT        1    550
FT   REPEAT      551   1107
FT   NP_BIND     174    229       ATP (POTENTIAL).
FT   NP_BIND     321    371       ATP (POTENTIAL).
SQ   SEQUENCE   1118 AA;  123914 MW;  887FAAE00AC07674 CRC64;

  Query Match           12.9%;  Score 82;  DB 1;  Length 1118;
  Best Local Similarity  28.3%;  Pred. No. 1.93e+00;
  Matches   15;  Conservative   16;  Mismatches  20;  Indels   2;  Gaps   2;

Db    702 DSIDVDQPEWSELTSVEEAKLFASKV-N-YPVLIRPSYVLSGAAMSVVNNEEE 752
          |  ::  ::::|  ::|| |:| :   : |||::    |||   |   :::|
Qy     35 DVNNIVTANRNEELTEEFYKVFVNEACHPYPVILPDRSVLSGDFTSAYADDDE 87

Search completed: Fri Aug 25 17:53:31 2000
Job time : 27 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 25 17:55:10 2000; MasPar time 3.23 Seconds
422.475 Million cell updates/sec

Tabular output not generated

Title: >US-09-257-585-4
Description: (1-89) from US09257585.pep
Perfect Score: 634
Sequence: 1 MKVQATFATLIALAAYFPAA.....DRSVLSGDFTSAYADDDESC 89

Scoring table: PAM 150
Gap 11

Searched: 152433 seqs, 15329240 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:6_COMB 4:PCT_COMB 5:backfiles1

Statistics: Mean 26.514; Variance 111.234; scale 0.238

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	78	12.3	286	1	US-08-118-	Sequence 65, Applicati	4.73e+01
2	78	12.3	286	4	PCT_US93-0	Sequence 65, Applicati	4.73e+01
3	77	12.1	412	1	US-07-712	Sequence 3, Applicatio	5.63e+01
4	76	12.0	200	2	US-08-836-	Sequence 4, Applicatio	6.70e+01
5	76	12.0	231	2	US 08 902	Sequence 6, Applicatio	6.70e+01
6	76	12.0	296	3	US-09-116-	Sequence 6, Applicatio	6.70e+01
7	76	12.0	296	1	US 08 597	Sequence 6, Applicatio	6.70e+01
8	76	12.0	529	2	US-08-836-	Sequence 3, Applicatio	6.70e+01
9	75	11.8	227	1	US-08-458-	Sequence 4, Applicatio	7.96e+01
10	75	11.8	493	2	US-08-933-	Sequence 2, Applicatio	7.96e+01
11	75	11.8	493	3	US-08-960-	Sequence 2, Applicatio	7.96e+01
12	75	11.8	1245	1	US-08-159	Sequence 8, Applicatio	7.96e+01
13	75	11.8	1245	1	US-08-304	Sequence 8, Applicatio	7.96e+01
14	75	11.8	1245	2	US-08-611-	Sequence 8, Applicatio	7.96e+01
15	75	11.8	1245	3	US 09 173	Sequence 8, Applicatio	7.96e+01
16	74	11.7	228	3	US-09-143-	Sequence 2, Applicatio	9.45e+01
17	74	11.7	262	1	US-08-392-	Sequence 37, Applicati	9.45e+01
18	74	11.7	262	3	US-09-330	Sequence 37, Applicati	9.45e+01
19	73	11.5	214	5	5254584-3	atent No. 5254584	1.12e+02
20	73	11.5	223	1	US-07-921-	Sequence 24, Applicati	1.12e+02
21	73	11.5	223	1	US-08-265	Sequence 24, Applicati	1.12e+02
22	73	11.5	223	1	PCT-US94-1	Sequence 24, Applicati	1.12e+02
23	73	11.5	223	1	US-07-745	Sequence 24, Applicati	1.12e+02
24	73	11.5	223	2	US-09-212-	Sequence 4,	
25	73	11.5	223	2	US-08-871-	Sequence 4,	
26	73	11.5	225	1	US-08-290-	Sequence 8,	
27	73	11.5	255	4	PCT-US94-0	Sequence 2,	
28	73	11.5	307	1	US 08 74b-	Sequence 14	
29	73	11.5	307	1	US 08 597-	Sequence 14	
30	73	11.5	320	1	US-08-461-	Sequence 10	
31	73	11.5	320	1	US-07 945-	Sequence 10	
32	73	11.5	320	1	US 08 461	Sequence 10	
33	73	11.5	320	1	US 08 462	Sequence 10	
34	73	11.5	320	4	PCT-US93 0	Sequence 10	
35	73	11.5	536	1	US-08-426-	Sequence 2,	
36	73	11.5	536	1	US-08 454-	Sequence 2,	
37	73	11.5	536	1	US 08 401-	Sequence 2,	
38	73	11.5	536	1	US 07 999-	Sequence 2,	
39	73	11.5	564	4	PCT-US93 0	Sequence 2,	
40	73	11.5	2161	2	US-08-223-	Sequence 4,	
41	73	11.5	2161	1	US 08-455-	Sequence 4,	
42	73	11.5	2161	1	US-08-455-	Sequence 5,	
43	73	11.5	2161	1	US 07 745	Sequence 2,	
44	73	11.5	2161	2	US 08 511	Sequence 2	
45	73	11.5	2161	2	US-08-223-	Sequence 5,	

ALIGNMENTS

RESULT 1
ID US-08-118-270-65 STANDARD; PRT; 286 AA
XX
AC xxxxxx
XX
DT
XX
DE Sequence 65, Application US/08118270
XX
CC Sequence 65, Application US/08118270
CC Patent No. 5508384
CC GENERAL INFORMATION:
CC APPLICANT: Murphy, Randall B.
CC APPLICANT: Schuster, David I.
CC TITLE OF INVENTION: POLYPEPTIDES OF G-PROTEI
CC TITLE OF INVENTION: RECEPTORS, AND COMPOSITI
CC NUMBER OF SEQUENCES: 148
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: BROWDY AND NEIMARK
CC STREET: 419 Seventh Street, N.W., Suite
CC CITY: Washington
CC STATE: D.C.
CC COUNTRY: USA
CC ZIP: 20004
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/118,270
CC FILING DATE: 09-SEP-1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/943,236
CC FILING DATE: 10-SEP-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Townsend, Kevin G.
CC REGISTRATION NUMBER: 34,033
CC REFERENCE/DOCKET NUMBER: MURPHY-2A
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202-628-5197
CC TELEFAX: 202-737-3528
CC TELEX: 248633
CC INFORMATION FOR SEQ ID NO: 65:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 286 amino acids
CC TYPE: amino acid

```
CC          STRANDEDNESS:  single
CC          TOPOLOGY:  linear
CC        MOLECULE TYPE:  peptide
SQ   SEQUENCE   286 AA; 31983 MW; 434446 CN;

  Query Match            12.4%;  Score 78;  DB 1;  Length 286;
  Best Local Similarity  33.3%;  Pred. No. 4.73e+01;
  Matches   14;  Conservative   11;  Mismatches  10;  Indels   5;  Gaps   4;

Db      89 LAVMAYDRYVA-ICHPLHYPVIVSSRIZVL-GSWAGGFG 125
           :::  |  :|   ||| |||:: |  || |  :::::
Qy      48 ITIEEYKVFVNEACHP--YPVILPDR-SVLSGDFTSAYA 83


RESULT    2
ID   PCT-US93-08528-65     STANDARD;     PRT;   286 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 65, Application PC/TUS9308528
XX
CC    Sequence 65, Application PC/TUS9308528
CC     GENERAL INFORMATION:
CC       APPLICANT:  New York University
CC       TITLE OF INVENTION:  POLYPEPTIDES OF G-COUPLED PROTEIN
CC       TITLE OF INVENTION:  RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
CC       NUMBER OF SEQUENCES:  348
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  BROWDY AND NEIMARK
CC         STREET:  419 Seventh Street, N.W., Suite 300
CC         CITY:  Washington
CC         STATE:  D.C.
CC         COUNTRY:  USA
CC         ZIP:  20004
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/US93/08528
CC         FILING DATE:  09-SEP-1993
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/943,236
CC         FILING DATE:  10-SEP-1992
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Townsend, Kevin G.
CC         REGISTRATION NUMBER:  34,033
CC         REFERENCE/DOCKET NUMBER:  MURPHY-2 PCT
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  202-628-5197
CC         TELEFAX:  202-737-3528
CC         TELEX:  248633
CC     INFORMATION FOR SEQ ID NO:  65:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  286 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  peptide
SQ   SEQUENCE   286 AA; 31983 MW; 434446 CN;

  Query Match            12.4%;  Score 78;  DB 4;  Length 286;
  Best Local Similarity  33.3%;  Pred. No. 4.73e+01;
  Matches   14;  Conservative   11;  Mismatches  10;  Indels   5;  Gaps   4;

Db      89 LAVMAYDRYVA-ICHPLHYPVIVSSRIZVL-GSWAGGFG 125
           :::  |  :|   ||| |||:: |  || |  :::::
Qy      48 ITIEEYKVFVNEACHP--YPVILPDR-SVLSGDFTSAYA 83


RESULT    3
ID   US-07-712-833A-3     STANDARD;     PRT;   412 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 3, Application US/07712833A
XX
CC    Sequence 3, Application US/07712833A
CC    Patent No. 5175101
CC     GENERAL INFORMATION:
CC       APPLICANT:  GOTZ, Friedrich
CC       APPLICANT:  SEEBER, Stefan
CC       TITLE OF INVENTION:  RECOMBINANT RESTRICTION ENZYME SAU3AI
CC       NUMBER OF SEQUENCES:  5
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Armstrong, Nikaido, Marmelstein, Kubovcik &
CC         ADDRESSEE:  Murray
CC         STREET:  1725 K Street N.W., Suite 1000
CC         CITY:  Washington D.C.
CC         COUNTRY:  United States of America
CC         ZIP:  20006
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/07/712,833A
CC         FILING DATE:  19910610
CC         CLASSIFICATION:  530
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  DE P 4018441.2
CC         FILING DATE:  08-JUN-1990
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Murray, Robert B.
CC         REGISTRATION NUMBER:  22,980
CC         REFERENCE/DOCKET NUMBER:  910808
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (202) 659-2930
CC         TELEFAX:  (202) 887-0357
CC         TELEX:  440142
CC     INFORMATION FOR SEQ ID NO:  3:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  412 amino acids
CC         TYPE:  AMINO ACID
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   412 AA; 47294 MW; 868600 CN;

  Query Match            12.1%;  Score 77;  DB 1;  Length 412;
  Best Local Similarity  29.5%;  Pred. No. 5.63e+01;
  Matches   13;  Conservative   12;  Mismatches  17;  Indels   2;  Gaps   2;

Db      47 DCYSKPFKNGIHSNKDIAQVSDFFMANIFADMIVGGFPCQDYSV 90
           ||   |:|:|   |: :  :|||:  |  ::|   :|: |:|
Qy      25 DCIIQRYKDG-DVNNIYTANRNEEITIEEYKVFVNE-ACHPYPV 66


RESULT    4
ID   US-08-836-442-4     STANDARD;     PRT;   200 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 4, Application US/08836442
XX
CC    Sequence 4, Application US/08836442
CC    Patent No. 5990293
CC     GENERAL INFORMATION:
CC       APPLICANT:  DOCHERTY, Andrew, J.P.
```

```
CC       APPLICANT:  SLOCOMBE, Patrick, M.
CC       TITLE OF INVENTION:  A HUMAN METALLOPROTEINASE
CC       TITLE OF INVENTION:  VARIANTS THEREOF AND DNA SEQUENCES CODING THEREFOR
CC       NUMBER OF SEQUENCES:  4
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
CC         STREET:  130 Water Street
CC         CITY:  Boston
CC         STATE:  MA
CC         COUNTRY:  USA
CC         ZIP:  02109
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette
CC         COMPUTER:  IBM Compatible
CC         OPERATING SYSTEM:  DOS
CC         SOFTWARE:  FastSEQ for Windows Version 2.0
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/846,442
CC         FILING DATE:  01-MAY-1997
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/IB97/02181
CC         FILING DATE:  13-MAR-1997
CC         APPLICATION NUMBER:  GB 9612150.4
CC         FILING DATE:  11-JUN-1996
CC         APPLICATION NUMBER:  GB 9526229.1
CC         FILING DATE:  21-DEC-1995
CC         APPLICATION NUMBER:  GB 9521498.7
CC         FILING DATE:  20-OCT-1995
CC         APPLICATION NUMBER:  GB 9521495.3
CC         FILING DATE:  20-OCT-1995
CC         APPLICATION NUMBER:  GB 9518023.8
CC         FILING DATE:  05-SEP-1995
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Resnick, David S.
CC         REGISTRATION NUMBER:  34,235
CC         REFERENCE/DOCKET NUMBER:  47425
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  617-523-3400
CC         TELEFAX:  617-523-6440
CC         TELEX:
CC     INFORMATION FOR SEQ ID NO:  4:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  200 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  No. 5990293e
SQ   SEQUENCE   200 AA; 23123 MW; 214496 CN;

  Query Match           12.0%;  Score 76;  DB 2;  Length 200;
  Best Local Similarity  27.5%;  Pred. No. 6.70e+01;
  Matches   14;  Conservative   12;  Mismatches  22;  Indels   3;  Gaps   3;

Db      80 WPGSVESPPRPHQAQLTAIPLACTFVC-L APMSTMPSPVSVGVVQDHS 128
           |:  ::  | |  |:  : ||      |:     |::  | ||:| :: |:|
Qy      23 WKDCIIQRYKDGDVNNIYTANRNEEITIEEYKVPVNEACHPYPV-ILPDRS 72


RESULT    5
ID   US-08-902-655A-6     STANDARD;     PRT;   231 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 6, Application US/08902655A
XX
CC    Sequence 6, Application US/08902655A
CC    Patent No. 5885819
CC     GENERAL INFORMATION:
CC       APPLICANT:  Kotod, Lene V.
CC       APPLICANT:  Kauppinen, Markus S.
CC       APPLICANT:  Christgau, Stephan
CC       APPLICANT:  Heldt-Hansen, Hans P.
CC       APPLICANT:  Dalboge, Henrik
CC       APPLICANT:  Andersen, Lene N.
CC       APPLICANT:  Si, Joan Q.
CC       APPLICANT:  Jacobson, Tina
CC       APPLICANT:  Munk, Niels
CC       APPLICANT:  Mullertz, Anette
CC       TITLE OF INVENTION:  ENZYMES WITH XYLANASE ACI
CC       TITLE OF INVENTION:  ASPERGILLUS ACULEATUS
CC       NUMBER OF SEQUENCES:  42
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  No. 5885819 No. 5885819disk of 7                    Arnt,
CC         STREET:  405 Lexington Avenue, 64th Floor
CC         CITY:  New York
CC         STATE:  New York
CC         COUNTRY:  United States of America
CC         ZIP:  10174-6401
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/902,655A
CC         FILING DATE:  30-July-1997
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Agris, Cheryl T.
CC         REGISTRATION NUMBER:  34,086
CC         REFERENCE/DOCKET NUMBER:  3954.214-US
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  212 867-1123
CC         TELEFAX:  212 878-9[illegible]
CC     INFORMATION FOR SEQ ID NO:  6:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  231 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   231 AA; 24944 MW; 211024 CN;

  Query Match           12.0%;  Score 76;  DB 2;  Length
  Best Local Similarity  42.9%;  Pred. No. 6.70e+01;
  Matches   11;  Conservative   [illegible];  Mismatches  12;

Db     150 YSDGSTYNIYTAIPINAPSQTAIP 179
           | ||   |||| | :  ::
Qy      31 YKDGDVNNIYTANRNEE[illegible]YKVP 56


RESULT    6
ID   US-09-116-622-6     STANDARD;     PRT;   [illegible] AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 6, Application US/09116622
XX
CC    Sequence 6, Application US/09116622
CC    Patent No. 6080[illegible]
CC     GENERAL INFORMATION:
CC       APPLICANT:  Kotod, Lene V.
CC       APPLICANT:  Kauppinen, Markus S.
CC       APPLICANT:  Christgau, Stephan
CC       APPLICANT:  Heldt-Hansen, Hans P.
CC       APPLICANT:  Dalboge, Henrik
CC       APPLICANT:  Andersen, Lene N.
CC       APPLICANT:  Si, Joan Q.
CC       APPLICANT:  Jacobson, Tina
CC       APPLICANT:  Munk, Niels
CC       APPLICANT:  Mullertz, Anette
```

```
CC       TITLE OF INVENTION:  ENZYMES WITH XYLANASE ACTIVITY FROM
CC       TITLE OF INVENTION:  ASPERGILLUS ACULEATUS
CC       NUMBER OF SEQUENCES:  42
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  No. 6080567o No. 6080567disk of No. 6080567th America, Inc.
CC         STREET:  405 Lexington Avenue, 64th Floor
CC         CITY:  New York
CC         STATE:  New York
CC         COUNTRY:  United States of America
CC         ZIP:  10174-6401
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/09/116,622
CC         FILING DATE:  16-July-1998
CC         CLASSIFICATION:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Aqris, Cheryl H.
CC         REGISTRATION NUMBER:  34,086
CC         REFERENCE/DOCKET NUMBER:  3954.224-US
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  212-867-0123
CC         TELEFAX:  212-878-9655
CC     INFORMATION FOR SEQ ID NO:  6:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  296 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   296 AA; 32379 MW; 496423 CN;

  Query Match           12.0%;  Score 76;  DB 3;  Length 296;
  Best Local Similarity  42.3%;  Pred. No. 6.70e+01;
  Matches   11;  Conservative    3;  Mismatches  12;  Indels   0,  Gaps   0;

Db     150 YSDGSTYNIYTATRTNAPSIQGTATF 175
           | ||   ||||| | :  :|:    |
Qy      31 YKDGDVNNIYTANRNEEITIEEYKVF 56


RESULT    7
ID   US-08-507-431-6      STANDARD;      PRT;   296 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 6, Application US/08507431
XX
CC    Sequence 6, Application US/08507431
CC    Patent No. 5693518
CC     GENERAL INFORMATION:
CC       APPLICANT:  Kolod, Lene V.
CC       APPLICANT:  Kauppinen, Markus S.
CC       APPLICANT:  Christgau, Stephan
CC       APPLICANT:  Heldt-Hansen, Hans P.
CC       APPLICANT:  Dalboge, Henrik
CC       APPLICANT:  Andersen, Lene N.
CC       APPLICANT:  Si, Joan Q.
CC       APPLICANT:  Jacobson, Tina
CC       APPLICANT:  Munk, Niels
CC       APPLICANT:  Mullertz, Anette
CC       TITLE OF INVENTION:  ENZYMES WITH XYLANASE ACTIVITY FROM
CC       TITLE OF INVENTION:  ASPERGILLUS ACULEATUS
CC       NUMBER OF SEQUENCES:  42
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  No. 5693518o No. 5693518disk of No. 5693518th America, Inc.
CC         STREET:  405 Lexington Avenue, 64th Floor
CC         CITY:  New York
CC         STATE:  New York
CC         COUNTRY:  United States of America
CC         ZIP:  10174-6401
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/507,431
CC         FILING DATE:  15-FEB-1996
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 60/002,800
CC         FILING DATE:  25-AUG-1996
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Harrington, James J.
CC         REGISTRATION NUMBER:  38,711
CC         REFERENCE/DOCKET NUMBER:  3954.204-US
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  212-867-0123
CC         TELEFAX:  212-878-9655
CC     INFORMATION FOR SEQ ID NO:  6:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  296 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   296 AA; 32379 MW; 496423 CN;

  Query Match           12.0%;  Score 76;  DB 1;  Length 296;
  Best Local Similarity  42.3%;  Pred. No. 6.70e+01;
  Matches   11;  Conservative    3;  Mismatches  12;  Indels   0;  Gaps   0;

Db     150 YSDGSTYNIYTATRTNAPSIQGTATF 175
           | ||   ||||| | :  :|:    |
Qy      31 YKDGDVNNIYTANRNEEITIEEYKVF 56


RESULT    8
ID   US-08-836-442-3      STANDARD;      PRT;   529 AA.
XX
AC   ^^^^^^
XX
DT
XX
DE   Sequence 3, Application US/08836442
XX
CC    Sequence 3, Application US/08836442
CC    Patent No. 5990293
CC     GENERAL INFORMATION:
CC       APPLICANT:  DOCHERTY, Andrew, J.P.
CC       APPLICANT:  SLOCOMBE, Patrick, M.
CC       TITLE OF INVENTION:  A HUMAN METALLOPROTEINASE
CC       TITLE OF INVENTION:  VARIANTS THEREOF AND DNA SEQUENCES CODING THEREFOR
CC       NUMBER OF SEQUENCES:  4
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
CC         STREET:  130 Water Street
CC         CITY:  Boston
CC         STATE:  MA
CC         COUNTRY:  USA
CC         ZIP:  02109
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette
CC         COMPUTER:  IBM Compatible
CC         OPERATING SYSTEM:  DOS
CC         SOFTWARE:  FastSEQ for Windows Version 2.0
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/836,442
CC         FILING DATE:  01-MAY-1997
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/GB96/02181
```

```
CC          FILING DATE:  14-MAR-1997
CC          APPLICATION NUMBER:  GB 9612150.4
CC          FILING DATE:  11-JUN-1996
CC          APPLICATION NUMBER:  GB 9526229.1
CC          FILING DATE:  21-DEC-1995
CC          APPLICATION NUMBER:  GB 9521498.7
CC          FILING DATE:  20-OCT-1995
CC          APPLICATION NUMBER:  GB 9521495.3
CC          FILING DATE:  20-OCT-1995
CC          APPLICATION NUMBER:  GB 9518023.8
CC          FILING DATE:  05-SEP-1995
CC        ATTORNEY/AGENT INFORMATION:
CC          NAME:  Resnick, David S
CC          REGISTRATION NUMBER:  34,235
CC          REFERENCE/DOCKET NUMBER:  47425
CC        TELECOMMUNICATION INFORMATION:
CC          TELEPHONE:  617-523-3400
CC          TELEFAX:  617-523-6440
CC          TELEX:
CC      INFORMATION FOR SEQ ID NO:  3:
CC        SEQUENCE CHARACTERISTICS:
CC          LENGTH:  529 amino acids
CC          TYPE:  amino acid
CC          STRANDEDNESS:  single
CC          TOPOLOGY:  linear
CC        MOLECULE TYPE:  No. 5990298e
SQ   SEQUENCE   529 AA; 60013 MW; 1383190 CN;

  Query Match           12.0%;  Score 76;  DB 2;  Length 529;
  Best Local Similarity  27.5%;  Pred. No. 6.70e+01;
  Matches   14;  Conservative   12;  Mismatches  22;  Indels   3;  Gaps   3;

Db    274 WRGSVISPRRKRHDIAQLLIATELAGTEVG-L-AFMSTMCSPYSVGVVQDRS 327
          |:  :: | |  |:  : ||      |:     |::  | ||:| :: |:|
Qy     23 WKDCIIQRYKDGDVNNIYTANRNEEIIEEYKVFVNEACHPYPV-ILPDRS 72


RESULT    9
ID   US-08-458-023B-4      STANDARD;      PRT;   227 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 4, Application US/08458023B
XX
CC    Sequence 4, Application US/08458023B
CC    Patent No. 5667990
CC     GENERAL INFORMATION:
CC       APPLICANT:  Berka, Randy M.
CC       APPLICANT:  Yoder, Wendy
CC       APPLICANT:  Takagi  Shinobu
CC       APPLICANT:  Boominathan, Karuppan C.
CC       TITLE OF INVENTION:  ASPERGILLUS EXPRESSION SYSTEM
CC       NUMBER OF SEQUENCES:  16
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  No. 5667990o No. 5667990disk of No. 5667990th America, Inc.
CC         STREET:  405 Lexington Avenue
CC         CITY:  New York
CC         STATE:  New York
CC         COUNTRY:  USA
CC         ZIP:  10174-6201
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/458,023B
CC         FILING DATE:  01-JUN-1995
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Lowney Dr., Karen A.
CC         REGISTRATION NUMBER:  31,274
CC         REFERENCE/DOCKET NUMBER:  4686.010 US
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  212-867-0123
CC         TELEFAX:  212 867-9555
CC     INFORMATION FOR SEQ ID NO:  4:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  227 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   227 AA; 25787 MW; 1387 75 CN;

  Query Match           11.8%;  Score 75;  DB 1;  Leng
  Best Local Similarity  40.8%;  Pred. No. 7.96e+01;
  Matches    8;  Conservative    8;  Mismatches  1 ;

Db    144 YTDGDQYDIFVSIRYNPESIDQIEIF 169
          | |||  :|: : | :: ;;  :|
Qy     31 YKDGDVNNIYTANRNEEIIEEYKVF 56


RESULT   10
ID   US-08-933-821-2      STANDARD;      PRT;   493 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application US/08933821
XX
CC    Sequence 2, Application US/08933821
CC    Patent No. 5972338
CC     GENERAL INFORMATION:
CC       APPLICANT:  Godowski, Paul ...
CC       APPLICANT:  Gurney, Austin L.
CC       TITLE OF INVENTION:  Tie Ligands
CC       NUMBER OF SEQUENCES:  17
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Genentech, Inc.
CC         STREET:  1 DNA Way
CC         CITY:  South San Francisco
CC         STATE:  California
CC         COUNTRY:  USA
CC         ZIP:  94080
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC DOS/MS-DOS
CC         SOFTWARE:  WinPatin (Genentech)
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/933,821
CC         FILING DATE:
CC         CLASSIFICATION:  530
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Dreger, Ginger R.
CC         REGISTRATION NUMBER:  33,055
CC         REFERENCE/DOCKET NUMBER:  P1140
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  650/225-3217
CC         TELEFAX:  650/952-9881
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  493 amino acids
CC         TYPE:  Amino Acid
CC         TOPOLOGY:  Linear
SQ   SEQUENCE   493 AA; 57134 MW; 1269880 CN;

  Query Match           11.8%;  Score 75;  DB 2;  Leng
  Best Local Similarity  34.2%;  Pred. No. 7.96e+01;
  Matches   13;  Conservative   12;  Mismatches  20;

Db    256 PIMPTLTSLPSSIDNRSGPWR-L-QALELDGHDISSIY 292
```

```
          :| :|| :|::    :: :|:||: |   || | ::||
Qy       5 ATFATLIALAAYFP-AANAWKDCIIQRYKDG-DVNNIY 40


RESULT   11
ID   US-08-960-507-2      STANDARD;      PRT;   493 AA.
XX
AC   XXXXXX
XX
DT
XX
DE   Sequence 2, Application US/08960507
XX
CC    Sequence 2, Application US/08960507
CC    Patent No. 6057435
CC     GENERAL INFORMATION:
CC       APPLICANT:  Godowski, Paul J
CC       APPLICANT:  Gurney, Austin L.
CC       TITLE OF INVENTION:  Tie Ligands
CC       NUMBER OF SEQUENCES:  24
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Genentech, Inc.
CC         STREET:  1 DNA Way
CC         CITY:  South San Francisco
CC         STATE:  California
CC         COUNTRY:  USA
CC         ZIP:  94080
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS DOS
CC         SOFTWARE:  WinPatin (Genentech)
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/960,507
CC         FILING DATE:
CC         CLASSIFICATION:  536
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Dreger, Ginger R.
CC         REGISTRATION NUMBER:  33,055
CC         REFERENCE/DOCKET NUMBER:  P1130p1
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  650/225-3216
CC         TELEFAX:  650/952-9881
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  493 amino acids
CC         TYPE:  Amino Acid
CC         TOPOLOGY:  Linear
SQ   SEQUENCE   493 AA; 57104 MW; 1263880 CN;

  Query Match            11.8%;  Score 75;  DB 3;  Length 493;
  Best Local Similarity  34.2%;  Pred. No. 7.96e+01;
  Matches    13;  Conservative    12;  Mismatches  10,  Indels    3,  Gaps    3,

Db     256 PTMPTLTSLPSSTDKPSCPWPDCL-QALEDGHDTSSIY 292
           :| :|| :|::    :: :|:||: |   || | ::||
Qy       5 ATFATLIALAAYFP AANAWKDCIIQRYKDG DVNNIY 40


RESULT   12
ID   US-08-158-232-8      STANDARD;      PRT;   1245 AA.
XX
AC   XXXXXX
XX
DT
XX
DE   Sequence 8, Application US/08158232
XX
CC    Sequence 8, Application US/08158232
CC    Patent No. 5596071
CC     GENERAL INFORMATION:
CC       APPLICANT:  Payne, Jewel
CC       APPLICANT:  Kennedy, M. Keith
CC       APPLICANT:  Randall, John Brooks
CC       APPLICANT:  Meier, Henry
CC       APPLICANT:  Uick, Heidi Jane
CC       APPLICANT:  Foncerrada, Luis
CC       APPLICANT:  Schnepf, H. Ernest
CC       APPLICANT:  Schwab, George E.
CC       APPLICANT:  Fu, Jenny
CC       TITLE OF INVENTION:  No. 5596071el Bacillus thuringiensis Toxins Active
CC       TITLE OF INVENTION:  Against Hymenopteran Pests
CC       NUMBER OF SEQUENCES:  51
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  David R. Saliwanchik
CC         STREET:  2421 N.W. 41st Street, Suite A-1
CC         CITY:  Gainesville
CC         STATE:  FL
CC         COUNTRY:  USA
CC         ZIP:  32606
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/158,232
CC         FILING DATE:
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/887,480
CC         FILING DATE:  22-MAY-1992
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/797,645
CC         FILING DATE:  25-NOV-1991
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/703,977
CC         FILING DATE:  22-MAY-1991
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Saliwanchik, David R.
CC         REGISTRATION NUMBER:  31,794
CC         REFERENCE/DOCKET NUMBER:  M/SCI104.C1
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  904-375-8100
CC         TELEFAX:  904-372-5800
CC     INFORMATION FOR SEQ ID NO:  8:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1245 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
CC       HYPOTHETICAL:  YES
CC       ANTI-SENSE:  NO
CC       ORIGINAL SOURCE:
CC         ORGANISM:  BACILLUS THURINGIENSIS
CC         INDIVIDUAL ISOLATE:  PS86Q3
CC       IMMEDIATE SOURCE:
CC         LIBRARY:  LAMBDAGEM (tm) - 11 library
CC         CLONE:  86Q3A
SQ   SEQUENCE   1245 AA; 139775 MW; 7983406 CN;

  Query Match            11.8%;  Score 75;  DB 1;  Length 1245;
  Best Local Similarity  28.6%;  Pred. No. 7.96e+01;
  Matches    10;  Conservative    11;  Mismatches  13;  Indels    1;  Gaps    1;

Db     388 VDSIVYSRKELQQLDIATYSTNNSNNCHPYGLRLS 422
           |::| |: :  ::: | |    :: |||| : |:
Qy      36 VNNI-YTANRNEEITIEEYKVFVNEACHPYPVILP 69


RESULT   13
ID   US-08-304-626-8      STANDARD;      PRT;   1245 AA.
XX
AC   XXXXXX
XX
```

```
DI
XX
DE   Sequence 8, Application US/08304626
XX
CC    Sequence 8, Application US/08304626
CC    Patent No. 5616495
CC     GENERAL INFORMATION:
CC       APPLICANT:  Payne, Jewel M.
CC       APPLICANT:  Kennedy, M. Keith
CC       APPLICANT:  Randall, John Brooks
CC       APPLICANT:  Meier, Henry
CC       APPLICANT:  Dick, Heidi Jane
CC       APPLICANT:  Foncerrada, Luis
CC       APPLICANT:  Schnepf, Harry E.
CC       APPLICANT:  Schwab, George E.
CC       TITLE OF INVENTION:  No. 5616495el Bacillus thuringiensis Isolates
CC       TITLE OF INVENTION:  Active Against Hymenopteran Pes s and Genes Encoding
CC       TITLE OF INVENTION:  Hymenopteran Active Toxins
CC       NUMBER OF SEQUENCES:  39
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  David R. Saliwanchik
CC         STREET:  2421 N.W. 41st Street, Suite A-1
CC         CITY:  Gainesville
CC         STATE:  FL
CC         COUNTRY:  USA
CC         ZIP:  32606
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/304,626
CC         FILING DATE:
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US/07/987,980
CC         FILING DATE:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Saliwanchik, David R.
CC         REGISTRATION NUMBER:  31,794
CC         REFERENCE/DOCKET NUMBER:  M/SC3104
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  904-375-8100
CC         TELEFAX:  904-372-5800
CC     INFORMATION FOR SEQ ID NO:  8:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1245 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
CC       HYPOTHETICAL:  YES
CC       ANTI-SENSE:  NO
CC       ORIGINAL SOURCE:
CC         ORGANISM:  BACILLUS THURINGIENSIS
CC         INDIVIDUAL ISOLATE:  PS86Q3
CC       IMMEDIATE SOURCE:
CC         LIBRARY:  LAMBDAGEM (tm) - 11 library
CC         CLONE:  86Q3A
SQ   SEQUENCE   1245 AA; 139775 MW; 7983406 CN;

  Query Match           11.8%;  Score 75;  DB 1;  Length 1245;
  Best Local Similarity  28.6%;  Pred. No. 7.96e+01;
  Matches   10;  Conservative   11;  Mismatches  13;  Indels   1,  Gaps   1;

Db    388 VDSIVYSPKEQQQLDIAIYSINNSNNCHPYCLRLS 422
          |:|| |: :  ::: |  |    :: |||| : |:
Qy     36 VNNI-YTANRNEEITIEEYKVFVNEACHPYPVILP 69


RESULT   14
ID   US-08-611-928-8     STANDARD;     PRT;   1245 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 8, Application US/ 8611928
XX
CC    Sequence 8, Application US/08611928
CC    Patent No. 5824792
CC     GENERAL INFORMATION:
CC       APPLICANT:  Payne, Jewel
CC       APPLICANT:  Kennedy, M. Keith
CC       APPLICANT:  Randall, John Brooks
CC       APPLICANT:  Meier, Henry
CC       APPLICANT:  Dick, Heidi Jane
CC       APPLICANT:  Foncerrada, Luis
CC       APPLICANT:  Schnepf, H. Ernest
CC       APPLICANT:  Schwab, George E.
CC       APPLICANT:  Fu, Jenny
CC       TITLE OF INVENTION:  No. 5824792el Bacillus th.                    s Isolates
CC       TITLE OF INVENTION:  Against Hymenopteran Pest
CC       NUMBER OF SEQUENCES:  39
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  David R. Saliwanchik
CC         STREET:  2421 N.W. 41st Street, Suite A-1
CC         CITY:  Gainesville
CC         STATE:  FL
CC         COUNTRY:  USA
CC         ZIP:  32606
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/611,928
CC         FILING DATE:  06-MAR-1996
CC         CLASSIFICATION:  530
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 08/158,232
CC         FILING DATE:  24-NOV-1993
CC         APPLICATION NUMBER:  US 07/887,980
CC         FILING DATE:  22-MAY-1992
CC         CLASSIFICATION:  530
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/797,645
CC         FILING DATE:  25-NOV-1991
CC         CLASSIFICATION:  530
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/703,977
CC         FILING DATE:  22-MAY-1991
CC         CLASSIFICATION:  530
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Saliwanchik, David R.
CC         REGISTRATION NUMBER:  31,794
CC         REFERENCE/DOCKET NUMBER:  M/SC3104.C1
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  904-375-8100
CC         TELEFAX:  904-372-5800
CC     INFORMATION FOR SEQ ID NO:  8:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1245 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
CC       HYPOTHETICAL:  YES
CC       ANTI-SENSE:  NO
CC       ORIGINAL SOURCE:
CC         ORGANISM:  BACILLUS THURINGIENSIS
CC         INDIVIDUAL ISOLATE:  PS86Q3
CC       IMMEDIATE SOURCE:
CC         LIBRARY:  LAMBDAGEM (tm)   11 library
```

```
CC         CLONE:  86Q3A
SQ   SEQUENCE   1245 AA; 139775 MW; 7983406 CN;

  Query Match           11.8%;  Score 75;  DB 2;  Length 1245;
  Best Local Similarity  28.6%;  Pred. No. 7.96e+01;
  Matches    10;  Conservative   11;  Mismatches  13;  Indels   1;  Gaps   1;

Db    388 VDSIVYSRKELQQLDIATYSTNNSNNCHPYCLKLS 422
          |::| |: :  ::: |  |    :: |||| : |:
Qy     36 VNNI-YTANRNEEITIEEYKVPVNEACHPYPVILP 69


RESULT   15
ID   US-09-173-891-8     STANDARD;     PRT;   1245 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 8, Application US/09173891
XX
CC    Sequence 8, Application US/09173891
CC    Patent No. 6077937
CC     GENERAL INFORMATION:
CC       APPLICANT:  Payne, Jewel
CC       APPLICANT:  Kennedy, M. Keith
CC       APPLICANT:  Randall, John Brooks
CC       APPLICANT:  Meier, Henry
CC       APPLICANT:  Uick, Heidi Jane
CC       APPLICANT:  Foncerrada, Luis
CC       APPLICANT:  Schnopf, H. Ernest
CC       APPLICANT:  Schwab, George E.
CC       APPLICANT:  Fu, Jenny
CC       TITLE OF INVENTION:  No. 6077937el Bacillus thuringiensis Toxins Active
CC       TITLE OF INVENTION:  Against Hymenopteran Pests
CC       NUMBER OF SEQUENCES:  51
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  David R. Saliwanchik
CC         STREET:  2421 N.W. 41st Street, Suite A-1
CC         CITY:  Gainesville
CC         STATE:  FL
CC         COUNTRY:  USA
CC         ZIP:  32606
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS DOS
CC         SOFTWARE:  Patentin Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/09/173,891
CC         FILING DATE:
CC         CLASSIFICATION:
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  08/158,232
CC         FILING DATE:
CC         APPLICATION NUMBER:  US 07/887,980
CC         FILING DATE:  22-MAY-1992
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/797,645
CC         FILING DATE:  25-NOV-1991
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/703,977
CC         FILING DATE:  22-MAY-1991
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Saliwanchik, David R.
CC         REGISTRATION NUMBER:  31,794
CC         REFERENCE/DOCKET NUMBER:  M/SCJ104.C1
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  904-375-8100
CC         TELEFAX:  904-372-5800
CC     INFORMATION FOR SEQ ID NO:  8:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1245 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
CC       HYPOTHETICAL:  YES
CC       ANTI-SENSE:  NO
CC       ORIGINAL SOURCE:
CC         ORGANISM:  BACILLUS THURINGIENSIS
CC         INDIVIDUAL ISOLATE:  PS86Q3
CC       IMMEDIATE SOURCE:
CC         LIBRARY:  LAMBDAGEM (tm) - 11 library
CC         CLONE:  86Q3A
SQ   SEQUENCE   1245 AA; 139775 MW; 7983406 CN;

  Query Match           11.8%;  Score 75;  DB 3;  Length 1245;
  Best Local Similarity  28.6%;  Pred. No. 7.96e+01;
  Matches    10;  Conservative   11;  Mismatches  13;  Indels   1;  Gaps   1;

Db    388 VDSIVYSRKELQQLDIATYSTNNSNNCHPYCLKLS 422
          |::| |: :  ::: |  |    :: |||| : |:
Qy     36 VNNI-YTANRNEEITIEEYKVPVNEACHPYPVILP 69


Search completed: Fri Aug 25 17:55:35 2000
Job time : 25 secs.
```

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 25 17:51:48 2000; MasPar time 4.33 Seconds
487.099 Million cell updates/sec

Tabular output not generated.

Title: >US-09-257-585-4
Description: (1-89) from US09257585.pep
Perfect Score: 634
Sequence: 1 MKVQATFATLIALAAYFPAA.....DRSVLSGDFTSAYADDDESC 89

Scoring table: PAM 150
Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq36
1:geneseqp

Statistics: Mean 28.189; Variance 114.078; scale 0.247

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	86	13.6	396	1	W23327	Meripilus giganteus xy	1.86e+01
2	82	12.9	315	1	W74867	Human secreted protein	3.83e+01
3	81	12.8	713	1	W05181	Neisseria gonorrhoeae	4.58e+01
4	81	12.8	1399	1	R38698	S-PPV-055 TCE virus gp	4.58e+01
5	81	12.8	1483	1	P80474	Sequence of protein of	4.58e+01
6	79	12.5	412	1	R15647	SamSAI methylase.	6.52e+01
7	79	12.5	713	1	W05180	Neisseria gonorrhoeae	6.52e+01
8	79	12.5	1099	1	W29683	Human Na-K-2Cl cotrans	6.52e+01
9	78	12.3	286	1	W02716	G protein coupled odor	7.78e+01
10	78	12.3	286	1	R48744	G protein coupled odor	7.78e+01
11	77	12.1	119	1	W98974	Alcaligenes sp. protei	9.26e+01
12	76	12.0	214	1	R25694	M-CSF isolated from hu	1.10e+02
13	76	12.0	306	1	R63791	Aspergillus aculeatus	1.10e+02
14	76	12.0	461	1	W74466	Adhesin/V cholerae tox	1.10e+02
15	76	12.0	530	1	W14772	Human metalloproteinas	1.10e+02
16	76	12.0	772	1	W14438	Type I carnitine palmi	1.10e+02
17	75	11.8	227	1	R75421	Humicola xylanase.	1.31e+02
18	75	11.8	227	1	R79231	Humicola insolens xyla	1.31e+02
19	75	11.8	493	1	Y05396	Human TIE ligand NL1 p	1.31e+02
20	75	11.8	495	1	W88482	Bovine papillomavirus	1.31e+02
21	75	11.8	495	1	P51099	Sequence of bovine pap	1.31e+02
22	75	11.8	993	1	R27650	Human calcium channel	1.31e+02
23	75	11.8	1245	1	R29029	Bacillus thuringiensis	1.31e+02
24	75	11.8	1245	1	W76[illegible]	Bacillus th	[illegible]
25	75	11.8	1245	1	W7[illegible]	B.t. toxin	[illegible]
26	75	11.8	1245	1	W13[illegible]	B6Q3a toxin	[illegible]
27	75	11.8	3993	1	[illegible]	Sequence of	[illegible]
28	74	11.7	402	1	W90[illegible]	Arabidopsis	[illegible]
29	73	11.5	164	1	R57[illegible]	Sequence of	[illegible]
30	73	11.5	218	1	W41[illegible]	Human colon	[illegible]
31	73	11.5	223	1	R75[illegible]	M-colony st	[illegible]
32	73	11.5	229	1	R23[illegible]	Human macro	[illegible]
33	73	11.5	233	1	R23[illegible]	Human macro	[illegible]
34	73	11.5	286	1	R30[illegible]	Subunit of	[illegible]
35	73	11.5	510	1	W10[illegible]	Neiderla	[illegible]
36	73	11.5	520	1	R12[illegible]	Human M-CS	[illegible]
37	73	11.5	535	1	W99[illegible]	Phaseolus	[illegible]
38	73	11.5	539	1	W93[illegible]	Human pCSF	[illegible]
39	73	11.5	536	1	R05[illegible]	Long form	[illegible]
40	73	11.5	546	1	P95[illegible]	Long form	[illegible]
41	73	11.5	554	1	R37[illegible]	Sequence of	[illegible]
42	73	11.5	554	1	R48[illegible]	Human MCSF	[illegible]
43	73	11.5	594	1	W72[illegible]	HSV-2 stra	[illegible]
44	73	11.5	662	1	W72[illegible]	HSV-2 stra	[illegible]
45	73	11.5	2161	1	W3[illegible]	Human calc	[illegible]

ALIGNMENTS

RESULT 1
ID W23327 standard; Protein; [illegible] AA.
AC W23327;
DT 14-JAN-1998 (first entry)
DE Meripilus giganteus xylanase.
KW Meripilus giganteus; Escherichia coli; xylanase.
KW agricultural waste; juice production; paper bleac[illegible]
KW filamentous fungi; alcohol; cell wall degradati[illegible]
OS Meripilus giganteus CBS 521.95.
PN WO9727290-A1.
PD 31-JUL-1997.
PF 22-JAN-1997; DK0030.
PR 22-JAN-1996; DK-000064.
PA (NOVO) NOVO-NORDISK AS.
PI Andersen LN, Dybdal L, Kruppinen MS, Kofod LV
DR WPI; 97-393678/36.
DR N-PSDB; T64848.
PT Meripilus giganteus CBS 521.95 DNA encoding a x[illegible]
PT providing enzyme for food preparation, animal f[illegible]
PT production and paper bleaching
PS Claim 14; Pages 49-40; 51pp; English.
CC This xylanase polypeptide sequence of 396 amino[illegible]
CC by a cDNA sequence from Meripilus giganteus str[illegible]
CC The xylanase encoding part of the DNA sequence [illegible]
CC plasmid pYES 2.0 present in E. coli DSM 10[illegible]
CC techniques. Host cells transformed with xylanase[illegible]
CC are cultured to produce this recombinant xylana[illegible]
CC Xylanases have plant cell wall degrading activi[illegible]
CC catalyse degradation of the backbone or side ch[illegible]
CC polysaccharide. The present xylanase enzyme [illegible]
CC inclusion in compositions, optionally with othe[illegible]
CC enzyme like x,[illegible] - particularly enzyme[illegible]
CC for degrading/modifying plant cell walls. The [illegible]
CC are useful in food preparation, production of [illegible]
CC products, separation of cereal (e.g. wheat [illegible]
CC wine or fruit juice production or as additives[illegible]
CC process animal feed. They can be used in beer[illegible]
CC modification of brewing by products to improve[illegible]
CC animal feed, and to reduce the viscosity of pl[illegible]
CC material. They are also useful in pulp or pape[illegible]
CC pulp bio-bleaching and in enzymatic breakdown[illegible]
CC wastes to produce alcohol fuels.
SQ Sequence 396 AA;

Query Match [illegible]%; Score 86; DB 1;
Best Local Similarity 4[illegible]%; Pred. No. 1.86e+01
Matches 11; Conservative 7; Mismatches [illegible];

```
Db      1 MKFSATFSALAALIPYALAQSPFWGQC 27
          ||  |||::| || :| :: :  | :|
Qy      1 MKVQATFATLIALAAY-FPAANAWKDC 26


RESULT    2
ID   W74867 standard; Protein; 315 AA.
AC   W74867;
DT   25-JAN-1999  (first entry)
DE   Human secreted protein encoded by gene 139 clone HSKGF03.
KW   Human; secreted protein; testis; tumour; foetal brain tissue;
KW   fusion protein; cancer; central nervous system; seizure;
KW   diagnosis; neurodegenerative disease.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Misc-difference 28
FT                   /label= unknown
FT   Misc-difference 315
FT                   /label= unknown
PN   WO9839448-A2.
PD   11-SEP-1998.
PF   06-MAR-1998; 00449 3.
PR   02-OCT-1997; US-061060.
PR   07-MAR-1997; US-038621.
PR   07-MAR-1997; US-040161.
PR   07-MAR-1997; US-040162.
PR   07-MAR-1997; US-040163.
PR   07-MAR-1997; US-040333.
PR   07-MAR-1997; US-040334.
PR   07-MAR-1997; US-040336.
PR   07-MAR-1997; US-040626.
PR   11-APR-1997; US-043311.
PR   11-APR-1997; US-043312.
PR   11-APR-1997; US-043313.
PR   11-APR-1997; US-043314.
PR   11-APR-1997; US-043568.
PR   11-APR-1997; US-043569.
PR   11-APR-1997; US-043576.
PR   11-APR-1997; US-043578.
PR   11-APR-1997; US-043580.
PR   11-APR-1997; US-043669.
PR   11-APR-1997; US-043670.
PR   11-APR-1997; US-043671.
PR   11-APR-1997; US-043672.
PR   11-APR-1997; US-043674.
PR   23-MAY-1997; US-047492.
PR   23-MAY-1997; US-047500.
PR   23-MAY-1997; US-047501.
PR   23-MAY-1997; US-047502.
PR   23-MAY-1997; US-047503.
PR   23-MAY-1997; US-047581.
PR   23-MAY-1997; US-047582.
PR   23-MAY-1997; US-047583.
PR   23-MAY-1997; US-047584.
PR   23-MAY-1997; US-047585.
PR   23-MAY-1997; US-047586.
PR   23-MAY-1997; US-047587.
PR   23-MAY-1997; US-047588.
PR   23-MAY-1997; US-047589.
PR   23-MAY-1997; US-047590.
PR   23-MAY-1997; US-047592.
PR   23-MAY-1997; US-047593.
PR   23-MAY-1997; US-047594.
PR   23-MAY-1997; US-047595.
PR   23-MAY-1997; US-047596.
PR   23-MAY-1997; US-047597.
PR   23-MAY-1997; US-047598.
PR   23-MAY-1997; US-047599.
PR   23-MAY-1997; US-047600.
PR   23-MAY-1997; US-047601.
PR   23-MAY-1997; US-047612.
PR   23-MAY-1997; US-047613.
PR   23-MAY-1997; US-047614.
PR   23-MAY-1997; US-047615.
PR   23-MAY-1997; US-047617.
PR   23-MAY-1997; US-047618.
PR   23-MAY-1997; US-047632.
PR   23-MAY-1997; US-047633.
PR   06-JUN-1997; US-048964.
PR   06-JUN-1997; US-048974.
PR   13-JUN-1997; US-049610.
PR   08-JUL-1997; US-051926.
PR   16-JUL-1997; US-052874.
PR   18-AUG-1997; US-055724.
PR   22-AUG-1997; US-056630.
PR   22-AUG-1997; US-056631.
PR   22-AUG-1997; US-056632.
PR   22-AUG-1997; US-056636.
PR   22-AUG-1997; US-056637.
PR   22-AUG-1997; US-056662.
PR   22-AUG-1997; US-056664.
PR   22-AUG-1997; US-056845.
PR   22-AUG-1997; US-056862.
PR   22-AUG-1997; US-056864.
PR   22-AUG-1997; US-056872.
PR   22-AUG-1997; US-056874.
PR   22-AUG-1997; US-056875.
PR   22-AUG-1997; US-056876.
PR   22-AUG-1997; US-056877.
PR   22-AUG-1997; US-056878.
PR   22-AUG-1997; US-056879.
PR   22-AUG-1997; US-056880.
PR   22-AUG-1997; US-056881.
PR   22-AUG-1997; US-056882.
PR   22-AUG-1997; US-056884.
PR   22-AUG-1997; US-056886.
PR   22-AUG-1997; US-056887.
PR   22-AUG-1997; US-056888.
PR   22-AUG-1997; US-056889.
PR   22-AUG-1997; US-056892.
PR   22-AUG-1997; US-056893.
PR   22-AUG-1997; US-056894.
PR   22-AUG-1997; US-056903.
PR   22-AUG-1997; US-056908.
PR   22-AUG-1997; US-056909.
PR   22-AUG-1997; US-056910.
PR   22-AUG-1997; US-056911.
PR   05-SEP-1997; US-057650.
PR   05-SEP-1997; US-057669.
PR   05-SEP-1997; US-057761.
PR   12-SEP-1997; US-058785.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA,
PI   Feng P, Ferrie AM, Fischer CL, Florence KA, Greene JM, Hu JS,
PI   Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA,
PI   Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
DR   WPI; 98-506364/43.
DR   N-PSDB; V59649.
PT   New isolated human genes and the secreted polypeptide(s) they encode
PT     useful for diagnosis and treatment of e.g. cancers, neurological
PT   disorders, immune diseases, inflammation or blood disorders
PS   Claim 1; Page 623-624; 721pp; English.
CC   This sequence represents a secreted human protein encoded by the nucleic
CC   acid molecule designated Gene 139 from the human cDNA clone HSKGF03
CC   (deposited as clone ATCC 97902 and ATCC 209048).
CC   The gene can be used to generate fusion proteins by linking to the gene
CC   to a human immunoglobulin Fc portion (e.g. V59502) for increasing the
CC   stability of the fused protein as compared to the human protein only.
CC   The invention relates to 186 novel genes and their fragments (nucleic
CC   acid sequences: V59511-V59812; amino acid sequences W74731-W75026) which
CC   are useful for preventing, treating or ameliorating medical conditions
CC   e.g. by protein or gene therapy. Also, pathological conditions can be
CC   diagnosed by determining the amount of the new polypeptides in a sample
CC   or by determining the presence of mutations in the new polynucleotides.
CC   Specific uses are described for each of the 186 polynucleotides, based on
```

```
CC   which tissues they are most highly expressed in (see V53611 for described
CC   uses).
SQ   Sequence   315 AA;

  Query Match           12.9%;  Score 82;  DB 1;  Length 315;
  Best Local Similarity  24.0%;  Pred. No. 3.88e-01;
  Matches   12;  Conservative   16;  Mismatches  20;  Indels   2;  Gaps   2;

Db    187 TLTASPGVFPVTVQPPHCVPDTYSNATIWYKIFTTAPDANTKYAQDYNPF 236
          || | :: || :    |: : | :: :  :|:|: |: :   ::|: |
Qy      9 TLIALAAYFPAANAWKDCIIQRYKDGDV-NNIYTANRNEEIT-IEEYKVF 56


RESULT    3
ID   W05181 standard; Protein; 713 AA.
AC   W05181;
DT   02-MAR-1998  (first entry)
DE   Neisseria gonorrhoeae strain FA1090 FrpB protein.
KW   FrpB; outer-membrane protein; iron regulating; immunogen;
KW   vaccine; diagnosis; infection; Neisseria gonorrhoeae;
KW   Neisseria meningitidis.
OS   Neisseria gonorrhoeae.
FH   Key             Location/Qualifiers
FT   Protein         1. .714
FT                   /label= frpB
FT                   /note= "70 kD major iron-regulated outer-membrane
FT                         protein"
PN   WO9631618-A1.
PD   10-OCT-1996.
PF   08-APR-1996; 004774.
PR   07-APR-1995; US-418964.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Beucher M, Sparling PF;
DR   WPI; 96-465048/46.
DR   N-PSDB; T42128.
PT   DNA encoding Neisseria iron-regulated outer membrane protein FrpB -
PT   useful for producing vaccine to protect mammals against N.
PT   gonorrhoeae and N. meningitidis infection
PS   Claim 1; Fig 10; 68pp; English.
CC   This sequence represents a novel iron-regulated outer-membrane protein
CC   FrpB which was isolated from the gonococcal microorganism Neisseria
CC   gonorrhoeae strain FA1090. This gene is also homologous to one found in
CC   Neisseria meningitidis. The FrpB protein is surface exposed and
CC   immunogenic and is predicted to be useful as a vaccine because of its
CC   surface exposure, partial antigenic conservation and susceptibility to
CC   attack by bacterial antibodies. Such a vaccine will be of use against
CC   infection in mammals by N. gonorrhoeae and N. meningitidis. The protein
CC   and its fragments will also be useful in the diagnosis of Neisseria
CC   type infections.
SQ   Sequence   713 AA;

  Query Match           12.8%;  Score 81;  DB 1;  Length 713;
  Best Local Similarity  32.8%;  Pred. No. 4.58e+01;
  Matches   19;  Conservative   15;  Mismatches  19;  Indels   5;  Gaps   4;

Db    186 KEGNFDGLFSYNRNDEKDYEAGKGFRNVNGGFTVFYSA-LLKRSYLA-KICTTFCDCD 241
          |:|: : ::: |||:.  | | |  || :  ||: | || |:  : ::::| |
Qy     42 KDKDVNNIYTANRNFPIIIFYYVF--VNFA-CHPYPVIIPFRSVLSGDTSAYADC 96


RESULT    4
ID   R38698 standard; Protein; 1399 AA.
AC   R38698;
DT   25-NOV-1993  (first entry)
DE   S-PRV-055 TGE virus qp195 gene product.
KW   Attenuated; vaccine; herpes virus; non-primate; live; safer; IBR;
KW   infectious bovine rhinotracheitis; MDV; Marek's disease virus; fowl;
KW   pseudo-rabies; swine; transmissible gastroenteritis virus.
OS   Transmissible gastroenteritis virus.
PN   US5223424-A.
PD   29-JUN-1993.
PF   27-JUL-1988; 225042.
PR   06-SEP-1985; US-773430.
PR   27-JAN-1986; US-823102.
PR   17-JUL-1986; US-887140.
PR   02-SEP-1986; US-902807.
PR   20-NOV-1986; US-933107.
PR   27-JUL-1987; US-078519.
PR   27-JUL-1988; US-225042.
PA   (PRUT-) PRUTECH RES & DEV.
PI   Chiang CH, Cochran MD, Macdonald RD;
DR   WPI, 93-215985/27.
DR   N-PSDB; Q42755.
PT   Recombinant fusion proteins for vaccine - comprises
PT   sequences fused to viral sequences e.g. pseudo-rab
PT   as vaccines
PS   Disclosure; Fig 22; 127pp; English.
CC   The sequence is that encoded by the transmissible
CC   virus qp195 gene from S PRV 055 which may be used
CC   of attenuated non-primate herpes viruses. These ca
CC   vaccines and provide a safer vaccine than current
CC   e.g. pseudorabies virus of swine, infectious bovin
CC   (IBR) virus or Marek's disease of fowl.
SQ   Sequence   1399 AA;

  Query Match           12.8%;  Score 81;  DB 1;  Leng
  Best Local Similarity  41.2%;  Prod. No. 4.58e+01;
  Matches    7;  Conservative    5;  Mismatches   5;

Db     41 VVLGDYFPTVQPWFNC 57
          : |: |||:  :| :|
Qy     11 IALAAYFPAANAWKIC 27


RESULT    5
ID   P80474 standard; protein; 1483 AA.
AC   P80474;
DT   19-NOV-1990  (first entry)
DE   Sequence of protein of transmissible gastroenteri
DE   on peplomer gene
KW   Pig vaccine.
OS   Transmissible gastroenteritis virus Purdue strain
FH   Key             Location/Qualifiers
FT   peptide         5. .19
FT                   /note= "Signal peptide"
FT   region          1394. .1414
FT                   /note= "Hydrophobic region which p
FT                   as a transmembrane anchor"
PN   EP-278541-A.
PD   17-AUG-1988.
PF   14-JAN-1988; 200004.
PR   16-JAN-1987; NL-000
PA   (DUIN) Duphar Int Res BV.
PI   Jacobs CE, Spaan WJM;
DR   WPI, 88-229403/33.
DR   N-PSDB; n81025.
PT   Transmissible gastroenteritis virus gene and prote
PT   obtd. by recombinant methods and used for prepara
PT   plus against transmissible gastro enteritis
PS   Claim 2; Fig 1; 12pp; English.
CC   A vaccine, an expression vector for the protein e
CC   transformed host cells are also new. The protein
CC   coupled to a carrier and used as vaccines for imm
CC   transmissible gastroenteritis (TGE). X in seq. i
CC   a stop codon in n81025.
SQ   Sequence   1483 AA;

  Query Match           12.8%;  Score 81;  DB 1;  Le
  Best Local Similarity  41.2%;  Prod. No. 4.58e+01;
  Matches    7;  Conservative    5;  Mismatches   5;

Db     56 VVLGDYFPTVQPWFNC 72
          : |: |||:  :| :|
Qy     11 IALAAYFPAANAWKIC 27
```

RESULT 6
ID R15647 standard; Protein; 412 AA.
AC R15647;
DT 16-MAR-1992 (first entry)
DE Sau3AI methylase.
KW Recombinant.
OS Staphylococcus carnosus.
PN EP-460673-A.
PD 11-DEC-1991.
PF 06-JUN-1991; 109302.
PR 08-JUN-1990; DE-018441.
PA (BOEF) BOEHRINGER MANNHEIM GMBH.
PI Gotz F, Seeber S;
DR WPI; 91-363167/50.
DR N-PSDB; Q15130.
PT New DNA encoding Sau3AI methylase and restriction enzymes
PT useful in recombinant DNA technology, and corresp. recombinant
PT vectors and transformed microorganisms
PS Disclosure; Page 14; 24pp; German.
CC The amino acid sequence codes for a protein with Sau3AI methylase
CC activity. Cloning of the Sau3AI restriction/methylase system is now
CC possible by using a host organism closely related to Staphylococcus
CC aureus 3A but lacking the system itself. See also R15503.
SQ Sequence 412 AA;

Query Match 12.5%; Score 79; DB 1; Length 412;
Best Local Similarity 29.5%; Pred. No. 6.52e+01;
Matches 13; Conservative 13; Mismatches 16; Indels 2; Gaps 2;

Db 47 DCYSKFPFKNGIHSNKHIGEVSDEEMANTEADMIVGGFPCQDYSV 90
|| |:|:| :| :: :|||:: | ::| :|: |:|
Qy 25 DCIIQPYKDGDVNNIYTAN-ENEEITIEEYKVFVNE-ACHPYPV 66

RESULT 7
ID W05180 standard; Protein; 713 AA.
AC W05180;
DT 27-FEB-1998 (first entry)
DE Neisseria gonorrhoeae strain FA19 FrpB protein.
KW FrpB; outer-membrane protein; iron regulating; immunogen;
KW vaccine, diagnosis, infection, Neisseria gonorrhoeae;
KW Neisseria meningitidis.
OS Neisseria gonorrhoeae.
FH Key Location/Qualifiers
FT Peptide 1. .23
FT Protein 24. .713
FT /label= FrpB
FT /note= "70 kD major iron-regulated outer-membrane
FT protein"
FT Region 32. .36
FT /label= TonB_box
PN WO9631618-A1.
PD 10-OCT-1996.
PF 08-APR-1996; 004774.
PR 07-APR-1995; US-418964.
PA (UYNC-) UNIV NORTH CAROLINA.
PI Beucher M, Sparling PF;
DR WPI; 96-465038/46.
DR N-PSDB; I42127.
PT DNA encoding Neisseria iron-regulated outer membrane protein FrpB
PT useful for producing vaccine to protect mammals against N.
PT gonorrhoeae and N. meningitidis infection
PS Claim 1; Fig 3; 68pp; English.
CC This sequence represents a novel iron-regulated outer-membrane protein
CC FrpB which was isolated from the gonococcal microorganism Neisseria
CC gonorrhoeae strain FA19. This gene is also homologous to one found in
CC Neisseria meningitidis. The FprB protein is surface exposed and
CC immunogenic and is predicted to be useful as a vaccine because of its
CC surface exposure, partial antigenic conservation and susceptibility to
CC attack by bacterial antibodies. Such a vaccine will be of use against
CC infection in mammals by N. gonorrhoeae and N. meningitidis. The protein
CC and its fragments will also be useful in the diagnosis of Neisseria
CC type infections.
SQ Sequence 713 AA;

Query Match 12.5%; Score 79; DB 1; Length 713;
Best Local Similarity 29.8%; Pred. No. 6.52e+01;
Matches 17; Conservative 15; Mismatches 22; Indels 3; Gaps 3;

Db 186 KPSNFDGLFSYNRNDFKDYFAGKGFFNDNGGKTVFYSALDFESYLA-KIGTTPGLGD 241
|:|: : ::: |||:| | | | |: : | | || |: : :::: |
Qy 32 KIGDVNNIYTANRNEEITIEEYKVFVNE-ACHPYP-VTLPDRSVLSGDFTSAYADDD 86

RESULT 8
ID W29683 standard; Protein; 1099 AA.
AC W29683;
DT 21-DEC-1998 (first entry)
DE Human Na K 2Cl cotransporter NKCC2.
KW Na-K-2Cl cotransporter; NaKCC2 gene; human; ion transport;
KW Bartter's syndrome; hypokalaemic alkalosis; hypercalciuria;
KW nephrocalcinosis; diagnosis; therapy.
OS Homo sapiens.
PN WO9829431-A1.
PD 09-JUL-1998.
PF 19-DEC-1997; U23553.
PR 31-DEC-1996; US-778052.
PA (UYYA) UNIV YALE.
PI Lifton RP, Simon DB;
DR WPI; 98-388029/33.
DR N-PSDB; V40562.
PT Thiazide sensitive cotransporter, ATP sensitive K channel and NaK2Cl
PT cotransporter genes - useful for developing products for the
PT diagnosis and treatment of ion transport disorders, e.g. Gitelman's
PT Syndrome or Bartter's Syndrome
PS Disclosure; Fig 7; 105pp; English.
CC This is the amino acid sequence of human renal bumetanide-sensitive
CC Na-K-2Cl cotransporter NKCC2 of the thick ascending limb of Henle's
CC loop. Mutations in the NKCC2 gene (see V40562) cause Bartter's
CC syndrome, a disorder featuring salt wasting and hypokalaemic
CC alkalosis associated with marked hypercalciuria and frequently
CC nephrocalcinosis. Different molecular variants inferred to alter
CC the structure of NKCC2 protein were identified in patients with
CC Bartter's syndrome by SSCP using 27 pairs of specific primers (see
CC V45677-730). Identification of the molecular basis of Bartter's
CC syndrome provides for the genetic diagnosis of this disorder. A
CC method is claimed for determining the presence or absence of a
CC mutated protein conferring altered ion transport by analysing
CC NKCC2 samples. Also claimed are vectors, host cells, a method of
CC producing NKCC2 protein, and an antibody that binds altered, but
CC not wild-type, NKCC2 protein. The products and methods can be used
CC for diagnosis of Bartter's syndrome, to identify carriers and to
CC identify modulators of NKCC2 function of therapeutic value.
SQ Sequence 1099 AA;

Query Match 12.5%; Score 79; DB 1; Length 1099;
Best Local Similarity 35.5%; Pred. No. 6.52e+01;
Matches 11; Conservative 7; Mismatches 12; Indels 1; Gaps 1;

Db 922 LTLLIPYILTERKKWKDCKLRIYVGGKINRI 952
| | :|: : : |||| :: | | :| |
Qy 10 LIALAAYFPAA-NAWKDCIIQRYKDGDVNNI 39

RESULT 9
ID W02716 standard; peptide; 286 AA.
AC W02716;
DT 13-NOV-1996 (first entry)
DE G-protein coupled odorant receptor I7.
KW G protein coupled receptor; ligand binding assay; transmembrane domain;
KW schizophrenia; dopamine; cAMP; adenosine; thrombin; adrenergic; opsin;
KW muscarinic acetylicholine, endothelin, bombesin; endocrine; rhodopsin;
KW odorant; cytomegalovirus; serotonergic.
OS Synthetic.
FH Key Location/Qualifiers
FT misc_difference 9

```
FT                     /note= "unknown amino acid"
PN   US5508384-A.
PD   16-APR-1996.
PF   10-SEP-1992; 943236.
PR   10-SEP-1992; US-943236.
PR   09-SEP-1993; US-118270.
PA   (UYNY ) UNIV NEW YORK STATE.
PI   Murphy RB,  Schuster DI;
DR   WPI; 96-208785/21.
PT   New dopamine receptor peptide - useful as antipsychotic agent, e.g.
PT   for treating schizophrenia
PS   Disclosure; Column 153-156; 184pp; English.
CC   Proteins W02657-W02730 represent a range of G protein coupled receptor
CC   (GPR) proteins selected from cAMP, adenosine, muscarinic acetylcholine,
CC   adrenergic, thrombin, endothelin, bombesin, endocrine, rhodopsin, opsin,
CC   odorant, cytomegaloviral and other GPR proteins.  The receptor proteins
CC   were used to design polypeptides, pref. based on the transmembrane
CC   domains, for use in G protein coupled receptor ligand binding assays.
CC   The polypeptide fragments retain biological activity such as binding a
CC   GPR ligand or modulating GPR ligand binding to a GPR (see W02747-W02999
CC   for examples of polypeptide fragments).  The polypeptide fragments can
CC   be used in compositions for treating subjects suffering from a pathology
CC   related to a GPR abnormality e.g. a  psychotic disorder such as
CC   schizophrenia.
SQ   Sequence   286 AA;

  Query Match           12.3%;  Score 78;  DB 1;  Length 286;
  Best Local Similarity  33.3%;  Pred. No. 7.78e+01;
  Matches   13;  Conservative   11;  Mismatches  19;  Indels   5;  Gaps   1;

Db      89 LAVMAYDRYVA-ICHPLHYPVIVSSRIZVL-GSWAGGFG 125
           ::: |  :|   ||| |||||: |  || |  :::::
Qy      48 IIIEEYKVFVNEACHP  YPVILPDR-SVLSGDFTSAYA 84


RESULT   10
ID   R48744 standard; Protein; 286 AA.
AC   R48744;
DT   07-JUN-1996  (first entry)
DE   G-protein coupled odorant receptor 17 protein.
KW   G-protein coupled receptor; ligand binding assay; transmembrane domain;
KW   psychotic disorder; schizophrenia; dopamine; cAMP; adenosine; thrombin;
KW   muscarinic acetylcholine; adrenergic; endothelin; bombesin; endocrine;
KW   rhodopsin; opsin; odorant; cytomegalovirus.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_difference 9
FT                   */note= "unknown amino acid"
PN   WO9405695-A1.
PD   17-MAR-1994.
PF   09-SEP-1993; 008528.
PR   10-SEP-1992; US-943236.
PA   (UYNY ) UNIV NEW YORK STATE.
PI   Murphy RB,  Schuster DI;
DR   WPI; 94-101120/12.
PT   Polypeptides of G coupled receptor proteins (GPRs) - useful for
PT   binding GPR ligands or modulating GPR binding
PS   Disclosure; Page 120; 160pp; English.
CC   Proteins R48685-R48758 represent a range of G-protein coupled receptor
CC   proteins selected from cAMP, adenosine, muscarinic acetylcholine,
CC   adrenergic, thrombin, endothelin, bombesin, endocrine, rhodopsin, opsin,
CC   odorant, cytomegaloviral and other G-protein coupled receptors.  The
CC   receptor proteins were used to design polypeptides, pref. based on the
CC   transmembrane domains, for use in G-protein coupled receptor ligand
CC   binding assays.  The polypeptide fragments retain biological activity
CC   such as binding a GPR ligand or modulating GPR ligand binding to a GPR
CC   (see R48759-R48798, R50569-R50807 and R89189-R89195 for examples of
CC   polypeptide fragments).  The polypeptide fragments can be used in
CC   compositions for treating subjects suffering from a pathology related to
CC   a GPR abnormality e.g. a psychotic disorder such as schizophrenia.
SQ   Sequence   286 AA;

  Query Match           12.3%;  Score 78;  DB 1;  Length 286;
  Best Local Similarity  33.3%;  Pred. No. 7.78e+01;
  Matches   13;  Conservative   11;  Mismatches  19;

Db      89 LAVMAYDRYVA-ICHPLHYPVIVSSRIZVL-GSWAGGFG 125
           ::: |  :|   ||| |||||: |  || |  :::::
Qy      48 IIIEEYKVFVNEACHP  YPVILPDR-SVLSGDFTSAYA 84


RESULT   11
ID   W98974 standard; Protein; 119 AA.
AC   W98974;
DT   10-MAY-1999  (first entry)
DE   Alcaligenes sp. protein PoxE.
KW   Alcaligenes, PoaR, PoaA, PoaB, PoaC; PoxD; PoxE; P
KW   PoxI; aromatic; oxidative; petroleum purification;
KW   drug.
OS   Alcaligenes sp.
PN   JP11042088-A.
PD   16-FEB-1999.
PF   25-JUL-1997; 200625.
PR   25-JUL-1997; JP-200625.
PA   (TOPU ) TONEN CORP.
DR   WPI; 99-197826/17.
DR   N-PSDB; X18867.
PT   New encoding an aromatic cpd. oxidative decompsn.
PT   the fields of petroleum purification, chemical and
PS   Claim 1; Page 7-19; 45pp; Japanese.
CC   The present sequence encodes Alcaligenes sp. prote
CC   Invention also describes PoaA, PoaB, PoaC, PoaD, P
CC   and PoxI. The proteins are useful in the fields o
CC   purification, chemical industry and drug industry,
CC   synthesis, conversion and decomposition of aromat
SQ   Sequence   119 AA;

  Query Match           12.1%;  Score 77;  DB 1;  Len
  Best Local Similarity  26.4%;  Pred. No. 9.26e+01;
  Matches   11;  Conservative   16;  Mismatches  27;

Db      10 RFEPADRCAVIFHGNKIYVGWLGELLAPHCFPFPFAMRMPEVV
           |:  :| : :: :||   :  :: :     | ||| :  | |:
Qy      39 KYFPCPVNNIYIANPNEIIIIIYKVFVNEACHPYPVILPDRSV

RESULT   12
ID   R25694 standard; protein; 214 AA.
AC   R25694;
DT   22-JAN-1993  (first entry)
DE   M-CSF isolated from human urine.
KW   osteopetrosis; Albers-Schonberg disease; ivory bo
KW   monocyte-macrophage colony stimulating factor; M
OS   Homo sapiens.
PN   JP04178434-A.
PD   25-JUN-1992.
PF   09-NOV-1990; 304538.
PR   09-NOV-1990; JP-304538.
PA   (MORC ) MORINAGA MILK IND CO LTD.
DR   WPI; 92-263009/32.
PT   New drug for treating osteopetrosis  contains hu
PT   monocyte-macrophage colony stimulating factor
PS   Claim 4(b); Page 1-2; 9pp; Japanese.
CC   A drug for osteopetrosis contains human monocyte-
CC   stimulating factor (M-CSF).  The M-CSF is obtaine
CC   a culture of human M-CSF-producing cells or a cult
CC   transformant cells expressing human M-CSF gene.
CC   sequence of human M-CSF isolated from human urine
CC   R25694 and of human M-CSF isolated from recombina
CC   in R25695.  The drug stimulates the proliferation
CC   and has fewer side effects than the present proli
CC   agent for osteoclasts.
SQ   Sequence   214 AA;

  Query Match           12.0%;  Score 76;  DB 1;  Len
  Best Local Similarity  27.1%;  Pred. No. 1.10e+02;
```

```
  Matches    12;  Conservative   17;  Mismatches  14;  Indels    3;  Gaps    3;

Db      58 PDT-MPDIMRPFDNTPNAIAIVQEQELSLRLKSCFTKDYEEHDRAC 102
           |:| :|: : | ::: :: ::: | : |: | : || | : | :|
Qy      46 EEITIEEYKVFVNEACHPYPVI-LDDRSV-LSGDFTSAYADDDESC 89


RESULT   13
ID   R63791 standard; Protein; 306 AA.
AC   R63791;
DT   07-JUN-1995  (first entry)
DE   Aspergillus aculeatus xylanase III .
KW   Xylanase III, Aspergillus aculeatus, alpha arabino pyranosidase, brewing,
KW   paper pulp; food preparation; plant cell wall degradation.
OS   Aspergillus aculeatus.
FH   Key             Location/Qualifiers
FT   misc_difference 300  306
FT                   /note= "no corresponding codons"
PN   WO9421785-A.
PD   29-SEP-1994.
PF   02 MAR 1994; DK0088.
PR   10-MAR-1993; DK-000268.
PR   14-OCT-1993; DK-001151
PA   (NOVO ) NOVO-NORDISK AS.
PI   Andersen LN,  Christgau S,  Dalboge H,  Heldt-hansen HP;
PI   Jacobsents,  Kauppinen MS,  Kofod LV,  Mullertz A;
PT   Munk N,  Si JQ;
DR   WPI; 94-317006/39
DR   N-PSDB; Q74638.
PT   New xylanase enzymes from Aspergillus aculeatus  used for
PT   degrading plant cell wall components, e.g. in the prepn. of feed,
PT   in baking and in prepn. of pulp or paper
PS   Claim 11; Page 57; 80pp; English.
CC   Q74638 encodes R63791 Aspergillus aculeatus xylanase II, which
CC   degrades plant cell wall components and reduces the viscosity
CC   of plant cell wall derived material. These properties are
CC   useful in the production of dough and baked products; in the
CC   preparation of feed, food, beer, wine, pulp and paper; and for the
CC   separation of cereal components. It can also be used in the
CC   production of antibodies.
SQ   Sequence   306 AA;

  Query Match           12.0%;  Score 76;  DB 1;  Length 306;
  Best Local Similarity  42.3%;  Pred. No. 1.1e+02;
  Matches    11;  Conservative    3;  Mismatches  12;  Indels    0;  Gaps    0;

Db     150 YSDGSTYNIYTATRENAPSIQGTATF 175
           | ||   |||| | :  :|:    |
Qy      31 YKDGDVNNIYTANRNEFITIEEYKVF 56


RESULT   14
ID   W74466 standard; Protein; 461 AA.
AC   W74466;
DT   18-MAY-1999  (first entry)
DE   Adhesin/V cholerae toxin A2 and B subunit fusion protein.
KW   Fusion gene; H. pylori infection; adhesin gene; V. cholerae toxin;
KW   A2 subunit; B subunit; diagnosis; vaccine; gastritis; gastric ulcer;
KW   duodenal ulcer.
OS   Chimeric - Helicobacter pylori.
OS   Chimeric - Vibrio cholerae.
PN   WO9853082-A1.
PD   26-NOV-1998.
PF   21-MAY-1997; 000091.
PR   21-MAY-1997; WO-KR0091.
PA   (DAEW-) DAEWOONG PHARM CO LTD.
PI   Kim BO, Lee BK, Park SK, Yoon SW, Yu YH;
DR   WPI; 98-434925/37
DR   N-PSDB; X21955.
PT   Fusion gene consisting of the Heli-bacter pylori adhesin gene
PT   ligated to the A2 and B subunit gene of Vibrio cholerae toxin -
PT   useful in the diagnosis and treatment of H. pylori
PS   Claim 2; Page 20-21; 32pp; English.
CC   This sequence represents the fusion protein of the invention, between the
CC   H. pylori adhesin gene, and the V. cholerae toxin A2 and B subunit genes.
CC   This protein is used in the diagnosis of and as a vaccine against
CC   H. pylori, which is the main cause of gastritis and gastric/duodenal
CC   ulcers. It may also be used in the production of an anti-H. pylori
CC   antibody. Vectors containing the DNA sequence and E. coli strains
CC   containing it form the basis of a process to produce the protein, which
CC   consists of culturing the microorganism and obtaining the protein. The
CC   protein is an effective vaccine against H. pylori due to its excellent
CC   immunogenicity for H. pylori, stability within the stomach environment,
CC   and its ability to penetrate through the stomach mucous membrane to
CC   stimulate sIgA production.
CC   Note: This sequence was indexed from WO9853082, which is the first
CC   major country equivalent to KR97059278.
SQ   Sequence   461 AA;

  Query Match           12.0%;  Score 76;  DB 1;  Length 461;
  Best Local Similarity  37.9%;  Pred. No. 1.1e+02;
  Matches    11;  Conservative    9;  Mismatches  7;  Indels    2;  Gaps    2;

Db       1 MK-KIAIAIAVALAGFASVAQAAWKKCLL 28
           ||  :::|  :|||:: : | | || |::
Qy       1 MKVQATFATLIALAAYFPAANA-WKDCIL 28


RESULT   15
ID   W14772 standard; Protein; 530 AA.
AC   W14772;
DT   19-JUN-1997  (first entry)
DE   Human metalloproteinase.
KW   Metalloproteinase; infectious disease; inflammation;
KW   autoimmune disease; allergy; atopic disease; cancer.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   domain          199. .399
FT                   /label= Catalytic_domain
PN   WO9709430-A2.
PD   13-MAR-1997.
PF   05-SEP-1996; G02181.
PR   05-SEP-1995; GB-018023.
PR   20-OCT-1995; GB-021498.
PR   20-OCT-1995; GB-021495.
PR   21-DEC-1995; GB-026229.
PR   11-JUN-1996; GB-012150.
PA   (CELL ) CELLTECH THERAPEUTICS LTD.
PI   Docherty AJP,  Slocombe PM;
DR   WPI; 97-192905/17.
DR   N-PSDB; T63123;
DR   N-PSDB; T63124.
PT   New isolated human metalloproteinase - used to develop prods. for
PT   the treatment of e.g. infectious disease, inflammatory disease,
PT   autoimmunity or allergic, atopic diseases
PS   Claim 4; Page 6; 16pp; English.
CC   A metalloproteinase (W14772) is the product of an isolated human
CC   DNA (T63123).  The enzyme, or its catalytic domain, can be produced
CC   in procaryotic or eukaryotic host cells and used to screen for
CC   cpds. that regulate its activity.  Such cpds. can be used to treat
CC   diseases or disorders in human in which the over- or unregulated
CC   prodn. of metalloproteinase is implicated.  The metalloproteinase
CC   can also be used to raise antibodies capable of regulating the
CC   activity of the enzyme.
SQ   Sequence   530 AA;

  Query Match           12.0%;  Score 76;  DB 1;  Length 530;
  Best Local Similarity  27.5%;  Pred. No. 1.1e+02;
  Matches    14,  Conservative   12,  Mismatches  22;  Indels    3;  Gaps    3;

Db     279 WPGSVLSPRERHDIAQLITATELASTIVG-L-AFMSTMGSFYSVGVVQNHS 327
           |:  :: | |  |:  : ||      |:     |::  | ||:| :: |:|
Qy      26 WKIGTIQGFKKGDAVNNIYIANSNLLPLPDAYEVNLACGMYPV IIDDRS 72


Search completed: Fri Aug 25 17:52:09 2000
```

Job time : 21 secs.

MPSRCH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 25 17:47:38 2000; MasPar time 4.22 Seconds
432.204 Million cell updates/sec

Tabular output not generated.

Title: >US-09-257-585-3
Description: (1-77) from US09257585.pep
Perfect Score: 568
Sequence: 1 MGPVPSLIWTTQFFANARQS RPCGLLSSSQCLERLHHPTL 77

Scoring table: PAM 150
Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq36
1:geneseqp

Statistics: Mean 27.792; Variance 102.012; scale 0.272

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	92	16.2	1621	1	W60162	Arabidopsis thaliana G	2.38e+00
2	91	16.0	381	1	W71015	Human neuroD2 protein,	2.92e+00
3	91	16.0	381	1	W22439	Human neurogenic diffe	2.92e+00
4	91	16.0	382	1	W22442	Mouse neurogenic diffe	2.92e+00
5	91	16.0	382	1	W71018	Murine neuroD2 protein	2.92e+00
6	90	15.8	667	1	W56133	Homo sapiens soluble K	3.58e+00
7	89	15.7	1622	1	W60163	Arabidopsis thaliana G	4.39e+00
8	82	14.4	322	1	R22209	h-IGFBP-4.	1.77e+01
9	82	14.4	424	1	R21833	Sequence of insulin-li	1.77e+01
10	82	14.4	1297	1	W40200	Infected cell protein	1.77e+01
11	82	14.4	1298	1	W80810	The amino acid sequenc	1.77e+01
12	79	13.9	830	1	W80314	CIITA (class II transa	3.19e+01
13	79	13.9	987	1	W57057	Class II trans activat	3.19e+01
14	79	13.9	1106	1	W80312	CIITA (class II transa	3.19e+01
15	79	13.9	1130	1	W57056	Class II trans activat	3.19e+01
16	79	13.9	1130	1	R81569	Class II transactivato	3.19e+01
17	79	13.9	1130	1	W80313	CIITA (class II transa	3.19e+01
18	79	13.9	1130	1	R72452	Class II transactivato	3.19e+01
19	79	13.9	1207	1	W80311	CIITA (class II transa	3.19e+01
20	78	13.7	238	1	R81453	Endothelial cell prote	3.87e+01
21	78	13.7	238	1	W74546	Amino acid sequence of	3.87e+01
22	78	13.7	1382	1	W21867	Human metastasis-asso	3.87e+01
23	77	13.6	55	1	Y07911	Human secreted protein	4.69e+01
24	77	13.6	1528	1	W33363	Human multi	[illegible]
25	77	13.6	2547	1	W81511	Drosophila	[illegible]
26	76	13.4	143	1	R38823	Alpha 6B in	[illegible]
27	76	13.4	1417	1	R96955	Multi-drug	[illegible]
28	76	13.4	1472	1	R96954	Multi-drug	[illegible]
29	76	13.4	1482	1	R96953	Multi-drug	[illegible]
30	76	13.4	1489	1	R96952	Multi-drug	[illegible]
31	76	13.4	1531	1	W99893	Human multi	[illegible]
32	76	13.4	1531	1	W99894	Human multi	[illegible]
33	76	13.4	1531	1	W74470	Human multi	[illegible]
34	76	13.4	1531	1	W74471	Human multi	[illegible]
35	76	13.4	1531	1	R93153	Multi drug	[illegible]
36	76	13.4	1531	1	R54928	Multidrug	[illegible]
37	76	13.4	1531	1	W57485	Human mult	[illegible]
38	76	13.4	1531	1	W57486	Human MRP	[illegible]
39	75	13.2	137	1	W21614	Streptococc	[illegible]
40	75	13.2	466	1	R71955	Erythroid	[illegible]
41	75	13.2	512	1	W96269	Chicken ac	[illegible]
42	75	13.2	1013	1	W40224	Human toll	[illegible]
43	75	13.2	1528	1	W99805	Mouse mult	[illegible]
44	75	13.2	1528	1	W74472	Mouse mult	[illegible]
45	75	13.2	1528	1	W57487	Murine mul	[illegible]

ALIGNMENTS

RESULT 1
ID W60162 standard; Protein; 1621 AA.
AC W60162;
DT 28-SEP-1998 (first entry)
DE Arabidopsis thaliana GS-X pump AtMRP2 polypeptide
KW AtMRP2 gene; GS-X pump; glutathione-S-conjugate
KW transgenic plant; bioremediation; herbicide
KW heavy metal tolerance; disease resistance; crop
OS Arabidopsis thaliana.
PN WO9821938-A1.
PD 28-MAY-1998.
PF 18-NOV-1997; 821436.
PR 08-OCT-1997; US-061328.
PR 18-NOV-1996; US-031040.
PA (UYPE-) UNIV PENNSYLVANIA
PI Li Z, Lu Y, Koa PA;
DR WPI; 98-312031/27.
DR N-PSDB; W34683.
PT New isolated plant glutathione S-conjugate pump
PT develop products for e.g. manipulating the
PT transport glutathionated compounds across cell
PS Example; Fig 15; 163pp; English.
CC This Arabidopsis thaliana polypeptide comprises
CC membrane transporter, the glutathione S-conjugate
CC AtMRP2, that uses MgATP as an energy source for
CC glutathionated derivatives of endogenous and exo
CC plants. Its amino acid sequence was deduced from
CC (see W34683), which was isolated from genomic DNA
CC homology to human MRP1. The invention also incl
CC and transgenic plants having an isolated DNA enc
CC pump and promoter sequences (see W34685-86) useful
CC the expression of other genes in a plant. The p
CC can be used to manipulate plants to affect heavy
CC herbicide detoxification, plant-pathogen interac
CC responses to (phyto)hormones, plant pigmentation,
CC and oxidative stress (all claimed). The plant G
CC serves to facilitate the vacuolar storage of ant
CC compounds induced following the hypersensitive
CC pathogens in the healthy cells surrounding lesio
CC Such a process is believed to limit the spread
CC limiting propagation of the pathogen and spatia
CC toxic action of the phytoalexin itself.
SQ Sequence 1621 AA;

Query Match 16.2%; Score 92; DB 1; L
Best Local Similarity 27.7%; Pred. No. 2.38e+0
Matches 16; Conservative 10; Mismatches 29;

```
Db    1374 PSVCQPQQLSLSPAIIPPSKEIVIDPATAAVDVPPDALIQKIIPEEPKSCTMLIIAHELN 1433
           |    ||  ||:  || :: :|   | |:::: |:       ||  :::| :|  :: |:
Qy      13 FFANAPQSTSLPPFLLFKNLLLAATE-TSTIE-RAHNESPSYIRHPYRPCGLLSSSQCLE 70


RESULT    2
ID   W71015 standard; Protein; 381 AA.
AC   W71015;
DT   21-OCT-1998  (first entry)
DE   Human neuroD2 protein, which is a bHLH protein.
KW   Basic helix loop helix, bHLH, neuroD; neuroectodermal tumour;
KW   classification; medulloblastoma; human.
OS   Homo sapiens.
PN   US5795723-A.
PD   18-AUG-1998.
PF   07-AUG-1997; 910973.
PR   07-AUG-1997; US-910973.
PR   06-MAY-1994; US-239238.
PR   02-NOV-1995; US-552142.
PA   (HUTC-) HUTCHINSON CANCER RES CENT FRED.
PI   Olson JM, Tapscott SJ;
DR   WPI; 98-466661/40.
DR   N-PSDB; V42931.
PT   Classifying neuroectodermal tumours from expression pattern of
PT   basic helix-loop-helix genes - especially for identifying
PT   medulla:blastoma and assessing its aggressiveness, specifically
PT   associated with expression of BHLH genes neuroD 1-3
PS   Example 11; Columns 59-62; 43pp; English.
CC   The present sequence represents a protein which is a member of the basic
CC   helix-loop-helix (bHLH) protein family, and is designated neuroD2. The
CC   bHLH genes are a family of genes associated with vertebrate neuronal,
CC   endocrinal and gastrointestinal development. The observed pattern of
CC   neuroD expression distinguishes subclasses of neuroectodermal tumours.
CC   The specification describes a method for the classification of human
CC   neuroectodermal tumours. The method comprises measuring, in a tumour
CC   sample, expression of at least one basic bHLH gene and identifying the
CC   tumour subclass by matching expression to predetermined expression
CC   profiles for known subclasses. For classifying the tumour as a
CC   medulloblastoma, the bHLH gene detected is neuroD1 and neuroD3.
CC   The method is used to classify neuroectodermal tumours, and to identify
CC   medullablastoma and for prognosis of this as aggressive.
SQ   Sequence   381 AA;

  Query Match           16.0%;  Score 91;  DB 1;  Length 381;
  Best Local Similarity  29.7%;  Pred. No. 2.92e+00;
  Matches   11;  Conservative   10;  Mismatches  15;  Indels   1;  Gaps   1;

Db    215 FLTEQGADGAGRFHGSGGPFAMHPYPYPGSRLAGAQC 251
          :|:   :    | |  : ::  ||| ||: |:::||
Qy     33 LLAATETSTIERAHNESPSYIRHPYR-PCGLLSSSQC 68


RESULT    3
ID   W22439 standard; Protein; 381 AA.
AC   W22439;
DT   02-OCT-1997  (first entry)
DE   Human neurogenic differentiation protein (NeuroD2).
KW   Neurogenic differentiation protein; NeuroD; neuroD2 gene;
KW   transcriptional activator; neuron; pancreas; gastrointestine;
KW   knock-out mouse; transgenic animal; cancer; diabetes; gene therapy.
OS   Homo sapiens
FH   Key             Location/Qualifiers
FT   region          137. .176
FT                   /label= HLH
FT                   /note= "helix-loop-helix region"
PN   WO9716548-A1.
PD   09-MAY-1997.
PF   30-OCT-1996; U17532.
PR   02-NOV-1995; US-552142.
PA   (HUTC-) HUTCHINSON CANCER RES CENT FRED.
PA   (WEIN/) WEINTRAUB N.
PI   Hollenberg SM,  Lee JE,  Tapscott SJ,  Weintraub HM;
DR   WPI; 97-272117/24.
DR   N-PSDB; T74890.
PT   Nucleic acid encoding neurogenic differentiation polypeptide -
PT   useful e.g. in regulating neuronal, endocrine and gastrointestinal
PT   development
PS   Claim 9; Page 62-63; 81pp; English.
CC   Human, mouse and Xenopus neurogenic differentiation proteins
CC   (W22439-43) are transcriptional activators involved in neuronal,
CC   endocrine and gastrointestinal development. They are active very
CC   early in embryo development and may represent an/off switches,
CC   possibly reversible, for controlling cell development. They were
CC   discovered by expression cloning and screening assays designed to
CC   identify basic-helix-loop-helix proteins capable of interacting
CC   with the protein product of the Drosophila daughterless gene. Human,
CC   mouse and frog neuroD gene sequences (T74887-94) have been
CC   identified. Cells expressing NeuroD polypeptides can be used to
CC   screen for anticancer agents able to drive terminal differentiation
CC   in neural tumours, to screen for (ant)agonists, or transplanted to a
CC   site of neuronal injury or disease, and possibly (since NeuroD can
CC   regulate insulin expression) to treat diabetics.
SQ   Sequence   381 AA;

  Query Match           16.0%;  Score 91;  DB 1;  Length 381;
  Best Local Similarity  29.7%;  Pred. No. 2.92e+00;
  Matches   11;  Conservative   10;  Mismatches  15;  Indels   1;  Gaps   1;

Db    215 FLTEQGADGAGRFHGSGGPFAMHPYPYPGSRLAGAQC 251
          :|:   :    | |  : ::  ||| ||: |:::||
Qy     33 LLAATETSTIERAHNESPSYIRHPYR-PCGLLSSSQC 68


RESULT    4
ID   W22442 standard; Protein; 382 AA.
AC   W22442;
DT   02-OCT-1997  (first entry)
DE   Mouse neurogenic differentiation protein (NeuroD2).
KW   Neurogenic differentiation protein; NeuroD; neuroD2 gene;
KW   transcriptional activator; neuron; pancreas; gastrointestine;
KW   knock-out mouse; transgenic animal; cancer; diabetes; gene therapy.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   region          138. .177
FT                   /label= HLH
FT                   /note= "helix-loop-helix region"
PN   WO9716548-A1.
PD   09-MAY-1997.
PF   30-OCT-1996; U17532.
PR   02-NOV-1995; US-552142.
PA   (HUTC-) HUTCHINSON CANCER RES CENT FRED.
PA   (WEIN/) WEINTRAUB N.
PI   Hollenberg SM,  Lee JE,  Tapscott SJ,  Weintraub HM;
DR   WPI; 97-272117/24.
DR   N-PSDB; T74893.
PT   Nucleic acid encoding neurogenic differentiation polypeptide -
PT   useful e.g. in regulating neuronal, endocrine and gastrointestinal
PT   development
PS   Claim 9; Page 70-71; 81pp; English.
CC   Human, mouse and Xenopus neurogenic differentiation proteins
CC   (W22439-43) are transcriptional activators involved in neuronal,
CC   endocrine and gastrointestinal development. They are active very
CC   early in embryo development and may represent an/off switches,
CC   possibly reversible, for controlling cell development. They were
CC   discovered by expression cloning and screening assays designed to
CC   identify basic-helix-loop-helix proteins capable of interacting
CC   with the protein product of the Drosophila daughterless gene. Human,
CC   mouse and frog neuroD gene sequences (T74887-94) have been
CC   identified. Cells expressing NeuroD polypeptides can be used to
CC   screen for anticancer agents able to drive terminal differentiation
CC   in neural tumours, to screen for (ant)agonists, or transplanted to a
CC   site of neuronal injury or disease, and possibly (since NeuroD can
CC   regulate insulin expression) to treat diabetics.
SQ   Sequence   382 AA;
```

```
  Query Match           16.0%;  Score 91;  DB 1;  Length 382;
  Best Local Similarity  29.7%;  Pred. No. 2.92e+00;
  Matches   11;  Conservative   10;  Mismatches  15;  Indels   1;  Gaps   1;

Db    216 FLTEQGADGAGRFHGSGGPFAMHPYPYPCSRLAGAQC 252
          :|:   :    | |  : ::  |||  ||: |:::|| 
Qy     33 LLAATETSTIERAHNESPSYIRHPYR-PGGLLSSSQC 68


RESULT    5
ID   W71018 standard; Protein; 382 AA.
AC   W71018;
DT   21-OCT-1998  (first entry)
DE   Murine neuroD2 protein, which is a bHLH protein.
KW   Basic helix-loop-helix; bHLH; neuroD; neuroectodermal tumour;
KW   classification; medulloblastoma; mouse.
OS   Mus musculus.
PN   US5795724-A.
PD   18-AUG-1998.
PF   07-AUG-1997; 910974.
PR   07-AUG-1997; US-910974.
PR   06-MAY-1994; US-239238.
PR   02-NOV-1995; US-552142.
PA   (HUTC ) HUTCHINSON CANCER RES CENT FRED.
PI   Olson JM, Tapscott SJ;
DR   WPI; 98-466661/40.
DR   N-PSDB; V42934.
PT   Classifying neuroectodermal tumours from expression pattern of
PT   basic-helix-loop-helix genes - especially for identifying
PT   medulla:blastoma and assessing its aggressiveness, specifically
PT   associated with expression of BHLH genes neuroD 1-3
PS   Example 11; Columns 71-74; 43pp; English.
CC   The present sequence represents a protein which is a member of the basic
CC   helix-loop-helix (bHLH) protein family, and is designated neuroD2. The
CC   bHLH genes are a family of genes associated with vertebrate neuronal,
CC   endocrinal and gastrointestinal development. The observed pattern of
CC   neuroD expression distinguishes subclasses of neuroectodermal tumours.
CC   The specification describes a method for the classification of human
CC   neuroectodermal tumours. The method comprises measuring, in a tumour
CC   sample, expression of at least one basic bHLH gene and identifying the
CC   tumour subclass by matching expression to predetermined expression
CC   profiles for known subclasses. For classifying the tumour as a
CC   medulloblastoma, the bHLH gene detected is neuroD1 and neuroD3.
CC   The method is used to classify neuroectodermal tumours, and to identify
CC   medullablastoma and for prognosis of this as aggressive.
SQ   Sequence   382 AA;

  Query Match           16.0%;  Score 91;  DB 1;  Length 382;
  Best Local Similarity  29.7%;  Pred. No. 2.92e+00;
  Matches   11;  Conservative   10;  Mismatches  15;  Indels   1;  Gaps   1;

Db    216 FLTEQGADGAGRFHGSGGPFAMHPYPYPCSRLAGAQC 252
          :|:   :    | |  : ::  |||  ||: |:::|| 
Qy     33 LLAATETSTIERAHNESPSYIRHPYR-PGGLLSSSQC 68


RESULT    6
ID   W56133 standard; Protein; 667 AA.
AC   W56133;
DT   17-AUG-1998  (first entry)
DE   Homo sapiens soluble KUZ protein.
KW   kuzbanian; kuz; neurogenic; KUZ protein; neuronal partitioning,
KW   development; NOTCH protein processing; regulation; cell function;
KW   signal transduction pathways; screening; receptor binding;
KW   metalloprotease; transmembrane; soluble.
OS   Homo sapiens.
PN   WO9808933-A1.
PD   05-MAR-1998.
PF   27-AUG-1997; U15099.
PR   23-JUL-1997; US-019390.
PR   29-AUG-1996; US-019390.
PA   (REGC ) UNIV CALIFORNIA.
PA   (UYYA ) UNIV YALE.
PI   Pan D, Rooke J, Kubli GM, Xu T, Yavari R;
DR   WPI, 98-177428/16.
DR   N-PSDB; V22649.
PT   New KUZ polypeptides, members of the ADAM family of
PT   useful in neural partitioning and development
PS   Claim 1; Pages 48-49; 58pp; English.
CC   The sequence is that of encoded by the kuzbanian ge
CC   gene. The soluble (lacking a transmembran domain)
CC   involved in neuronal partitioning and development
CC   in processing of the NotCH protein by cleaving the
CC   The KUZ protein can be used to screen for compound
CC   binding of KUZ to its receptor or the cleavage of
CC   protein, hence acting to regulate NoTCH signal tra
CC   pathways and regulate cell functions.
SQ   Sequence   667 AA;

  Query Match           15.8%;  Score 90;  DB 1;  Leng
  Best Local Similarity  28.2%;  Pred. No. 4.58e+00;
  Matches   13;  Conservative    7;  Mismatches  19;

Db    551 MRSIVSLLLWSLQCFKNEKLLEKLPALQFTLL 584
          | :: ||| |: | | | :   : |:  |  ||
Qy      1 MCPVFSLL-WTTQFLANARQSCTPPLLPFKNLL 33


RESULT    7
ID   W60163 standard; Protein; 1622 AA.
AC   W60163;
DT   28-SEP-1998  (first entry)
DE   Arabidopsis thaliana GS X pump AtMRP1 polypeptide.
KW   AtMRP1 gene; GS-X pump; glutathione-S-conjugate t
KW   transgenic plant; bioremediation; herbicide toleran
KW   heavy metal tolerance; disease resistance; GS-X pu
OS   Arabidopsis thaliana.
PN   WO9821938-A1.
PD   28-MAY-1998.
PF   18-NOV-1997; U21336.
PR   08-OCT-1997; US-061428.
PR   18-NOV-1996; US-031841.
PA   (GYPE ) UNIV PENNSYLVANIA.
PI   Li Z, Lu Y, Rea PA;
DR   WPI; 98-312041/27.
DR   N-PSDB; V34684.
PT   New isolated plant glutathione S conjugate pump DN
PT   develop products for e.g. manipulating the ability
PT   transport glutathionated compounds across cell mem
PS   Example; Fig 18; 184pp; English.
CC   This Arabidopsis thaliana polypeptide comprises a
CC   membrane transporter. The glutathione S-conjugate
CC   AtMRP1, that uses MgATP as an energy source for t
CC   glutathionated derivatives of endogenous and exoge
CC   plants. Its amino acid sequence was deduced from
CC   (see V34684), which was isolated from genomic DNA
CC   homology to human MRP1. The invention also includ
CC   and transgenic plants having an isolated DNA encod
CC   pump and promoter sequences (see V34685-86) useful
CC   the expression of other genes in a plant. The pr
CC   can be used to manipulate plants to affect heavy m
CC   herbicide detoxification, plant-pathogen interacti
CC   responses to (phyto)hormones, plant pigmentation,
CC   and oxidative stress (all claimed). The plant GS
CC   serves to facilitate the vacuolar storage of antib
CC   compounds induced following the hypersensitive res
CC   pathogens in the healthy cells surrounding primary
CC   Such a process is believed to limit the spread of
CC   limiting propagation of the pathogen and spatially
CC   toxic action of the phytoalexin itself.
SQ   Sequence   1622 AA;

  Query Match           15.7%;  Score 89;  DB 1;  Leng
  Best Local Similarity  26.7%;  Pred. No. 4.49e+00;
  Matches   16;  Conservative   16;  Mismatches  25;
```

```
Db    1374 FSVGQRQLLSLARALLRRSKLLVLDEATAAVDVRTDVLIQKTIREEFKSCTMLIIAHRLN 1433
           |    ||  ||:  || :: :|   | |:::: |:       ||  :::| :|  :: |:
Qy      13 FFANAPQSTSIPPFLLFKNLLLAATP-TSTIE-PAHNESPSYIRHPYPPCQLLSFSQCLE 70


RESULT    8
ID    R22209 standard; Protein; 322 AA.
AC    R22209;
DT    10-JUL-1992  (first entry)
DE    h-IGFBP-4.
KW    Insulin-like growth factor binding protein; cancer; osteoporosis;
KW    anaemia; abnormal growth; antibody.
OS    Homo sapiens.
PN    WO9203152-A.
PD    05-MAR-1992.
PF    28-AUG-1992; U06139.
PR    28-AUG-1990; US-574613.
PR    31-AUG-1990; US-576648.
PA    (CHIR-) CHIRON CORP.
PI    Kiefer MC, Masiarz F, Zapf JL, Born WH;
DR    WPI; 92-096592/12.
DR    N-PSDB; Q22964.
PT    New insulin-like growth factor binding protein IGFBP-4 - for
PT    diagnosis and treatment of cancer, osteoporosis, anaemia and
PT    abnormal growth
PS    Claim 2; Fig 1; 68pp; English.
CC    The protein sequence was deduced from the DNA sequence obtd. by
CC    screening a human osteosarcoma (Ost 3) cDNA library with BP4 probes
CC    IGF-BP4 is useful for treating osteoporosis, in association with
CC    bound IGF. The IGF/IGF BP combination is useful for preventing
CC    hypoglycaemia and possible mitogenic effects at injection sites.
CC    The binding protein potentiates the erythropoietin-like effect of
CC    IGF-1. It may also be used to target IGF to specific tissues.
CC    When administered alone IGF-BP4 may be used for blocking the
CC    adverse effects of IGFs and to control excessive growth in IGF-BP
CC    deficient subjects.
SQ    Sequence   322 AA;

  Query Match           14.4%;  Score 82;  DB 1;  Length 322;
  Best Local Similarity  50.0%;  Pred. No. 1.77e+01;
  Matches   11;  Conservative    6,  Mismatches   5,  Indels   0,  Gaps   0;

Db     28 LPPLLLLLALLLAASPGGALAR 49
          |||:||:  ||||||:  ::: |
Qy     23 LPPFLLFKNLLLAATETSTIER 44


RESULT    9
ID    R21833 standard; Protein; 324 AA.
AC    R21833;
DT    23-JUL-1992  (first entry)
DE    Sequence of insulin-like-growth-factor binding
DE    protein 4 (IGF BP-4).
KW    Wound healing; growth; erythropoiesis; chondrogenesis.
OS    Homo sapiens.
FH    Key             Location/Qualifiers
FT    peptide         22. .48
FT                    /label= signal
FT    protein         48. .262
PN    WO9203469-A.
PD    05-MAR-1992.
PF    28-AUG-1991; U06138.
PR    28-AUG-1990; US-574613.
PR    31-AUG-1990; US-576629.
PA    (CHIR-) CHIRON CORP.
PI    Kiefer MC, Masiarz F;
DR    WPI; 92-096829/12.
DR    N-PSDB; Q22674.
PI    DNA encoding insulin-like-growth-factor binding protein 4 -
PT    useful for high level protein production with applications in
PT    e.g. wound healing and growth
PS    Disclosure; Fig 1; 67pp; English.
CC    A human osteosarcoma (Ost 3) cDNA library was prepared, contg. 2.3
CC    times 10(7) recombinant clones, propagated in E. coli BB4 and
CC    blotted onto nitrocellulose filters. Screening was with BP4 probes,
CC    labelled with T4 polynucleotide kinase and [lambda32P]ATP.
CC    Bluescript SK(-) plasmids contg. BP4 cDNA inserts were used to
CC    obtain DNA inserts which were sequenced and one clone, pBSBP4.1
CC    encoded IGFBP-4. "X" in R21833 denotes translated stop codon.
SQ    Sequence   324 AA;

  Query Match           14.4%;  Score 82;  DB 1;  Length 324;
  Best Local Similarity  50.0%;  Pred. No. 1.77e+01;
  Matches   11;  Conservative    6,  Mismatches   5;  Indels   0;  Gaps   0;

Db     28 LPPLLLLLALLLAASPGGALAR 49
          |||:||:  ||||||:  ::: |
Qy     23 LPPFLLFKNLLLAATETSTIER 44


RESULT   10
ID    W40200 standard; Protein; 1297 AA.
AC    W40200;
DT    30-JUN-1998  (first entry)
DE    Infected cell protein number 4.
KW    Infected cell protein number 4; ICP4; alpha-4; cell apoptosis;
KW    therapeutics.
OS    Herpes simplex virus.
FH    Key             Location/Qualifiers
FT    Domain          100. .200
FT                    /note= "Transactivation domain"
FT    Domain          300. .500
FT                    /note= "DNA binding domain"
FT    Domain          700. .750
FT                    /note= "Nuclear localisation domain"
FT    Domain          750. .1298
FT                    /note= "Transactivation domain"
PN    WO9804709-A2.
PD    05-FEB-1998.
PF    23-JUL-1997; U12904.
PR    26-JUL-1996; US-690473.
PA    (ARCH-) ARCH DEV CORP.
PI    Leopardi R, Roizman B;
DR    WPI; 98-136097/12.
DR    N-PSDB; V10362.
PT    Use of herpes simplex virus ICP4 polypeptide  useful for, e.g.
PT    blocking apoptosis in cells, production of proteins and gene therapy
PS    Disclosure; Fig 2; 63pp; English.
CC    The herpes simplex virus (HSV) infected cell protein number 4
CC    (ICP4) is encoded by the alpha-4 gene. Both the alpha-4 gene and ICP4
CC    can be used to block cell apoptosis. Similarly the administration of an
CC    agent that inhibits ICP4 or the alpha-4 gene can induce apoptosis in HSV
CC    infected cells. This can be used for the immortalisation of cells,
CC    production of proteins, gene therapy, or inhibition of cell death induced
CC    in vivo. They can also be used for production of therapeutics comprising
CC    inhibitors of HSV ICP4 function, useful for treating HSV function.
SQ    Sequence  1297 AA;

  Query Match           14.4%;  Score 82;  DB 1;  Length 1297;
  Best Local Similarity  28.4%;  Pred. No. 1.77e+01;
  Matches   19;  Conservative   19;  Mismatches  25;  Indels   4,  Gaps   4;

Db    999 LLSTRDLAFAGAVEFLGLLASAGDRRLIVVNT-VRACDWP-ADGPAVSRQHAYLACELLP 1056
          || | :: || | :  :| :    : |::  |   : :::  ::|:  | |:| :| ||:
Qy      7 LLWTTQP-PANAPQSTSIPPFLLFKNLLLAATETSTIEPAHNESPSYIR HPYPPCQLLS 64

Db   1057 AVQCAVR 1063
          : ||  |
Qy     65 SSQCLER 71


RESULT   11
ID    W80810 standard; Protein; 1298 AA.
AC    W80810;
DT    29-JAN-1999  (first entry)
DE    The amino acid sequence of the Herpes simplex virus ICP4.
```

KW Herpes simplex virus ICP4; HSV ICP4; modulator; apoptosis;
KW stimulation; inhibition; HSV infection.
OS Herpes simplex virus.
PN WO9846637-A2.
PD 22-OCT-1998.
PF 16-APR-1998; 98WO-IB000573.
PR 16-APR-1997; 97US-0843659.
PA (ARCH-) ARCH DEV CORP.
PI Leopardi R, Roizman B;
DR WPI; 98-594559/50.
DR N-PSDB; V68520.
PT Use of herpes simplex virus U(s)3 polypeptide - for developing
PT products for modulating apoptosis in cells and for identifying
PT compounds which act as stimulators or inhibitors of apoptosis
PS Example 2; Pages 63-67; 85pp; English.
CC This is the amino acid sequence of Herpes simplex virus ICP4 used in
CC the method of the invention as modulators of apoptosis. The methods
CC and products can be used to identify compounds which modulate
CC (stimulate or inhibit) apoptosis in cells. They can be used to
CC immortalise cells for the study of these cells or for growing cells
CC in large numbers for the productions of proteins. They can also be
CC used for stimulating apoptosis in cells, e.g. for treating a subject
CC with a HSV infection.
SQ Sequence 1298 AA;

Query Match 14.4%; Score 82; DB 1; Length 1298;
Best Local Similarity 28.4%; Pred. No. 1.77e-01;
Matches 19; Conservative 19; Mismatches 25; Indels 4; Gaps 4;

Db 1000 LISTRDLAPAGAVFPTGLLASAGDPPLIVVNT-VPACDWP-ADGPAVSRQIAYLACELLP 1057
Qy 7 LLWTTQF-FANARQSTSLPPFLIFKNIILAATETSTIERAHNESPSYIP-HPYRPCGLLS 64

Db 1058 AVQCAVR 1064
Qy 65 SSQCLER 71

RESULT 12
ID W80314 standard; Protein; 830 AA.
AC W80314;
DT 01-FEB-1999 (first entry)
DE CIITA (class II transactivator) type IV protein.
KW CIITA gene; class II transactivator; MHC class II molecule;
KW interferon-gamma; interleukin-4; vaccine; cancer treatment.
OS Homo sapiens.
PN EP-874049-A1.
PD 28-OCT-1998.
PF 21-APR-1998; 400968.
PR 22-APR-1997; FR-004954.
PA (TRGE) TRANSGENE SA.
PI Mach B;
DR WPI; 98-559115/48.
PT DNA encoding MHC class II trans-activator polypeptide(s) - useful
PT e.g. as primers for enzymatic amplification, as detection probes or
PT as inhibitors of expression of genes encoding MHC in dendritic cells
PS Disclosure; Pages 66-69; 86pp; French.
CC The present sequence represents type IV CIITA (class II transactivator)
CC protein. The products can be used to treat diseases for which
CC enhanced expression of genes coding for MHC class II molecules is
CC desired, especially where the enhanced expression is desired in
CC dendritic cells or after induction by a cytokine, especially
CC interferon-gamma or interleukin-4. Inhibitors of the gene or protein
CC can be used to treat diseases for which reduced expression of genes
CC coding for MHC class II molecules is desired or can be used as vaccines,
CC especially for cancer treatment.
SQ Sequence 830 AA;

Query Match 13.9%; Score 79; DB 1; Length 830;
Best Local Similarity 29.4%; Pred. No. 3.19e+01;
Matches 12; Conservative 16; Mismatches 11; Indels 2; Gaps 2;

Db 473 LLGPSAAASVDRKQKVLARYLKP-LQP-GTLPARQLLELLH 511
Qy 33 LLAATETSTIERAHNESPSYIPHPYRPCGLLSSSQCLERLH 73

RESULT 13
ID W57057 standard; Protein; [illegible] AA.
AC W57057;
DT 17-AUG-1998 (first entry)
DE Class II trans activator (CIITA) 151 deletion mutant
KW Class II trans activator; MHC; CIITA; autoimmune d
KW transplantation; xenograft; major histocompatibil
KW gene therapy; arthritis; ribozyme; 151 deletion mu
OS Homo sapiens.
OS Synthetic.
FH Key Location/Qualifiers
FT Misc_difference 6
FT /note "wild type 143 amino acid
FT after this position."
FT Misc_difference 7
FT /label E1*01
FT /note "wild type Glu is replaced
FT Misc_difference 8
FT /label C1*01
FT /note "wild type Leu is replaced
PN WO9815626-A2.
PD 16-APR-1998.
PF 08-OCT-1997; 002751.
PR 21-MAR-1997; GB-005913.
PR 08-OCT-1996; GB-020644.
PA (CHIL-) INST CHILD HEALTH.
PI Fabre JW, Gustafsson KI, Yan S;
DR WPI; 98-249813/21.
DR N-PSDB; V28717.
PT Deleted form of the class II trans-activator that
PT antigen expression - and ribozyme directed to tra
PT related nucleic acid, vectors, transformed cells
PT used for treating auto immune disease and to inhi
PT rejection
PS Claim 15; Pages 69-72; 144pp; English.
CC This represents a 151 deletion mutant of the class
CC (CIITA) polypeptide. The N terminal region of the
CC deleted and can be used for reducing the expressi
CC histocompatibility complex (MHC) class II antigen
CC to bases 1159-1161 (GUA) of human CIITA mRNA, or
CC other species and the CIITA polypeptide can be use
CC of MHC class II antigens, particularly for treatme
CC disease (e.g. arthritis and diabetes) or to treat
CC intended as source of xenografts. The ribozyme and
CC may be generated in vivo by gene therapy, using th
CC acids targeted for localised suppression of the in
CC from transgenic animals in which at least some ce
CC transfected with the CIITA encoding nucleic acids
CC human transplantation. The CIITA polypeptide supp
CC class II antigens, in cells that express them con
CC lymphokine induction. It has no transcription act
CC regulatory proteins in the promoter region of cla
CC competitively inhibits endogenous CIITA.
SQ Sequence 987 AA;

Query Match 13.9%; Score 79; DB 1; Len
Best Local Similarity 29.4%; Pred. No. 3.19e+01;
Matches 12; Conservative 16; Mismatches 11;

Db 640 LLGPSAAASVDRKQKVLARYLKP-LQP-GTLPARQLLELLH 678
Qy 33 LLAATETSTIERAHNESPSYIPHPYRPCGLLSSSQCLERLH 73

RESULT 14
ID W80312 standard; Protein; 1106 AA.
AC W80312;
DT 01-FEB-1999 (first entry)
DE CIITA (class II transactivator) type II protein.

KW CIITA gene; class II transactivator; MHC class II molecule;
KW interferon-gamma; interleukin-4; vaccine; cancer treatment.
OS Homo sapiens.
PN EP-874049 A1.
PD 28-OCT-1998.
PF 21-APR-1998; 400968.
PR 22-APR-1997; FR-004954.
PA (TRGE) TRANSGENE SA.
PI Mach B;
DR WPI; 98-559115/48.
PT DNA encoding MHC class II trans-activator polypeptide(s) - useful
PT e.g. as primers for enzymatic amplification, as detection probes or
PT as inhibitors of expression of genes encoding MHC in dendritic cells
PS Claim 10; Pages 58-61; 86pp; French.
CC The present sequence represents type II CIITA (class II transactivator)
CC protein. The products can be used to treat diseases for which
CC enhanced expression of genes coding for MHC class II molecules is
CC desired, especially where the enhanced expression is desired in
CC dendritic cells or after induction by a cytokine, especially
CC interferon-gamma or interleukin-4. Inhibitors of the gene or protein
CC can be used to treat diseases for which reduced expression of genes
CC coding for MHC class II molecules is desired or can be used as vaccines,
CC especially for cancer treatment.
SQ Sequence 1106 AA;

Query Match 14.9%; Score 79; DB 1; Length 1106;
Best Local Similarity 29.3%; Pred. No. 3.19e+01;
Matches 12; Conservative 16; Mismatches 11; Indels 2; Gaps 2;

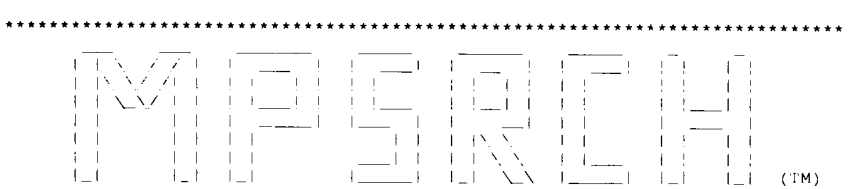
Db 749 LLGPSAAASVDRKQKVLARYLKR-LQP-GTLRARQLLELLH 787
Qy 33 LLAATETSTIERAHNESPSYIRHPYRPCGLLSSSQCLERLH 73

RESULT 15
ID W57056 standard; Protein; 1130 AA.
AC W57056;
DT 17-AUG-1998 (first entry)
DE Class II trans activator (CIITA) polypeptide.
KW Class II trans activator; MHC; CIITA; autoimmune disease; treatment;
KW transplantation; xenograft; major histocompatibility complex; diabetes;
KW gene therapy; arthritis; ribozyme.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Region 52. .137
FT /note= "N-terminal region rich in glutamate/aspartate"
FT Region 163. .195
FT /note= "region rich in proline/serine/threonine"
FT Region 209. .237
FT /note= "region rich in proline/serine/threonine"
FT Region 261. .422
FT /note= "region rich in proline/serine/threonine"
PN WO9815626-A2.
PD 16-APR-1998.
PF 08-OCT-1997; G02751.
PR 21-MAR-1997; GB-005911.
PR 08-OCT-1996; GB-020940.
PA (CHIL-) INST CHILD HEALTH.
PI Fabre JW, Gustafsson KI, Yun S;
DR WPI; 98-240813/21.
DR N-PSDB; V28716.
PT Deleted form of the class II trans-activator that reduces class II
PT antigen expression and ribozyme directed to trans-activator mRNA,
PT related nucleic acid, vectors, transformed cells and antibodies,
PT used for treating auto-immune disease and to inhibit xeno-graft
PT rejection
PS Claim 14; Fig 5; 104pp; English.
CC This represents a class II trans activator (CIITA) polypeptide. When
CC the N-terminal region of the CIITA is deleted, it can be used for
CC reducing the expression of major histocompatibility complex (MHC) class
CC II antigens. A ribozyme targeted to bases 1159-1161 (GUA) of human CIITA
CC mRNA, or corresponding target in other species and the CIITA polypeptide
CC can be used to reduce expression of MHC class II antigens, particularly
CC for treatment of autoimmune disease (e.g. arthritis and diabetes) or to
CC treat non human animals intended as source of xenografts. The ribozyme
CC and the CIITA polypeptide may be generated in vivo by gene therapy, using
CC the encoding nucleic acids targeted for localised suppression of the
CC immune response. Material from transgenic animals in which at least some
CC cells are stably transfected with the CIITA encoding nucleic acids are
CC used for animal to-human transplantation. The CIITA polypeptide
CC suppresses synthesis of class II antigens, in cells that express them
CC constitutively or after lymphokine induction. It has no transcription
CC activity but still binds to regulatory proteins in the promoter region of
CC class II genes, so competitively inhibits endogenous CIITA.
SQ Sequence 1130 AA;

Query Match 13.9%; Score 79; DB 1; Length 1130;
Best Local Similarity 29.3%; Pred. No. 3.19e+01;
Matches 12; Conservative 16; Mismatches 11; Indels 2; Gaps 2;

Db 773 LLGPSAAASVDRKQKVLARYLKR-LQP-GTLRARQLLELLH 811
Qy 33 LLAATETSTIERAHNESPSYIRHPYRPCGLLSSSQCLERLH 73

Search completed: Fri Aug 25 17:47:57 2000
Job time : 19 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 25 17:51:10 2000; MasPar time 3.03 Seconds
389.428 Million cell updates/sec

Tabular output not generated.

Title: >US-09-257-585-3
Description: (1-77) from US09257585.pep
Perfect Score: 568
Sequence: 1 MGPVPSLLWITQFFANARQS RPCGLLSSSQCLERLHHPTIL 77

Scoring table: PAM 150
Gap 11

Searched: 152433 seqs, 15329240 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:6_COMB 4:PCT_COMB 5:backfiles1

Statistics: Mean 26.140; Variance 101.525; scale 0.257

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	91	16.0	381	1	US-08-910-	Sequence 11, Applicati	2.14e+00
2	91	16.0	382	1	US-08-910-	Sequence 17, Applicati	2.14e+00
3	82	14.4	322	5	5212074-1	atent No. 5212074	1.23e+01
4	82	14.4	1298	2	US-08-690-	Sequence 2, Applicatio	1.23e+01
5	79	13.9	1130	2	US-08-519-	Sequence 6, Applicatio	2.18e+01
6	78	13.7	204	2	US-08-531-	Sequence 48, Applicati	2.63e+01
7	78	13.7	204	2	US-08-718-	Sequence 48, Applicati	2.63e+01
8	78	13.7	238	2	US-08-878-	Sequence 2, Application	2.63e+01
9	78	13.7	238	2	US-09-082	Sequence 2, Application	2.63e+01
10	78	13.7	238	1	US-08-289-	Sequence 2, Application	2.63e+01
11	78	13.7	238	1	US-08-884-	Sequence 2, Application	2.63e+01
12	76	13.4	141	1	US-08-241	Sequence 5, Application	3.83e+01
13	76	13.4	141	3	US-07-695-	Sequence 5, Application	3.83e+01
14	76	13.4	1531	2	US-08-460-	Sequence 2, Application	3.83e+01
15	76	13.4	1531	3	US-08-461-	Sequence 2, Application	3.83e+01
16	76	13.4	1531	3	US-08-407-	Sequence 2, Application	3.83e+01
17	76	13.4	1531	1	US-08-141-	Sequence 2, Application	3.83e+01
18	76	13.4	1531	3	US-08-463-	Sequence 2, Application	3.83e+01
19	76	13.4	1531	3	US-08-463-	Sequence 4, Application	3.83e+01
20	76	13.4	1531	2	US-08-462	Sequence 4, Application	3.83e+01
21	76	13.4	1531	2	US-08-462	Sequence 2, Application	3.83e+01
22	76	13.4	1531	1	US-08-463-	Sequence 2, Application	3.83e+01
23	76	13.4	1541	1	US-08-463-	Sequence 4, Application	3.83e+01
24	76	13.4	1541	2	US-08-460-	Sequence 4, Application	3.83e+01
25	76	13.4	1541	3	US-08-461-	Sequence 4, Application	3.83e+01
26	75	13.2	402	1	US-08-325-	Sequence 1, Application	4.99e+01
27	75	13.2	402	1	US-08-460-	Sequence 11, Application	4.99e+01
28	75	13.2	466	1	US-07-923	Sequence 2, Application	4.99e+01
29	75	13.2	1013	3	US-08-991	Sequence 2, Application	4.99e+01
30	75	13.2	1013	2	US-08-866-	Sequence 5, Application	4.99e+01
31	75	13.2	1013	2	US-08-921	Sequence 5, Application	4.99e+01
32	75	13.2	1528	2	US-08-462	Sequence 6, Application	4.99e+01
33	75	13.2	1528	1	US-08-463	Sequence 6, Application	4.99e+01
34	75	13.2	1528	3	US-08-463	Sequence 6, Application	4.99e+01
35	75	13.2	1528	3	US-08-461-	Sequence 6, Application	4.99e+01
36	75	13.2	1528	2	US-08-460-	Sequence 6, Application	4.99e+01
37	74	13.0	247	2	US-08-218-	Sequence 4, Application	6.51e+01
38	74	13.0	247	2	US-08-851-	Sequence 4, Application	6.51e+01
39	74	13.0	415	3	US-08-149-	Sequence 8, Application	8.51e+01
40	74	13.0	415	2	US-08-484	Sequence 8, Application	8.51e+01
41	74	13.0	415	2	US-08-484-	Sequence 8, Application	8.51e+01
42	74	13.0	415	3	US-08-458-	Sequence 8, Application	8.51e+01
43	74	13.0	415	2	US-08-484-	Sequence 8, Application	8.51e+01
44	74	13.0	415	2	US-08-480-	Sequence 8, Application	8.51e+01
45	73	12.9	169	2	US-08-455-	Sequence 45, Application	1.11e+02

ALIGNMENTS

RESULT 1
ID US-08-910-973-11 STANDARD; PRT; 381 AA.
XX
AC XXXXXX
XX
DT
XX
DE Sequence 11, Application US/08910973
XX
CC Sequence 11, Application US/08910973
CC Patent No. 5795723
CC GENERAL INFORMATION:
CC APPLICANT: Lapsett, Stephen J.
CC APPLICANT: Olson, James M.
CC TITLE OF INVENTION: Expression of Neurotrophic [illegible]
CC NUMBER OF SEQUENCES: 24
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Christensen, O'Connor Johnson, K
CC STREET: 1420 Fifth Avenue, Suite 2800
CC CITY: Seattle
CC STATE: WA
CC COUNTRY: USA
CC ZIP: 98101 2347
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC Compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/910,973
CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/08/239,238
CC FILING DATE: 06-MAY-1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/08/05741
CC FILING DATE: 06-MAY-1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US96/17542
CC FILING DATE: 30-October-1996
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Sheliness, Diana K.
CC REGISTRATION NUMBER: 35,456
CC REFERENCE/DOCKET NUMBER: FBCR-1 10958
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 206-682-8100; 206-224-0735 (fi

```
CC          TELEFAX:  206-225-0779
CC      INFORMATION FOR SEQ ID NO:  11:
CC        SEQUENCE CHARACTERISTICS:
CC          LENGTH:  381amino acids
CC          TYPE:  amino acid
CC          TOPOLOGY:  linear
CC        MOLECULE TYPE:  protein
CC        ORIGINAL SOURCE:
CC          ORGANISM:  Homo sapiens
SQ   SEQUENCE   381 AA; 41400 MW; 703930 CN;

  Query Match            16.0%;  Score 91;  DB 1;  Length 381;
  Best Local Similarity  29.7%;  Pred. No. 2.14e+00;
  Matches   11;  Conservative   10;  Mismatches  15;  Indels   1;  Gaps   1;

Db     215 FLTEQGADGACRFHGSGGPFAMHPYPYPGSRLAGAQC 251
           :|:   :    | |  : ::  |||  ||: |:::|| 
Qy      33 LLAATETSTIERAHNESPSYIRHPYR-PCGLLSSSQC 68


RESULT    2
ID   US-08-910-973 17      STANDARD;      PRT;   382 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 17, Application US/08910973
XX
CC    Sequence 17, Application US/08910973
CC    Patent No. 5795723
CC    GENERAL INFORMATION:
CC      APPLICANT:  Tapscott, Stephen J.
CC      APPLICANT:  Olson, James M.
CC      TITLE OF INVENTION:  Expression of Neurogenic bHLH Genes in Primitive Neurocyte
CC      NUMBER OF SEQUENCES:  24
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Christensen O'Connor Johnson KindnessPLLC
CC        STREET:  1420 Fifth Avenue, Suite 2800
CC        CITY:  Seattle
CC        STATE:  WA
CC        COUNTRY:  USA
CC        ZIP:  98101-2347
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/910,973
CC        FILING DATE:
CC        CLASSIFICATION:  435
CC      PRIOR APPLICATION DATA:
CC        APPLICATION NUMBER:  US 08/239,238
CC        FILING DATE:  06-MAY-1994
CC      PRIOR APPLICATION DATA:
CC        APPLICATION NUMBER:  WO PCT/US95/05741
CC        FILING DATE:  08-MAY-1995
CC      PRIOR APPLICATION DATA:
CC        APPLICATION NUMBER:  PCT/US95/12542
CC        FILING DATE:  30-October-1996
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Sheiness, Diana K.
CC        REGISTRATION NUMBER:  35,356
CC        REFERENCE/DOCKET NUMBER:  FHCR-1-10958
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  206-682-8100; 206 224-0735 (direct)
CC        TELEFAX:  206-225 0779
CC    INFORMATION FOR SEQ ID NO:  17:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  382 amino acids
CC        TYPE:  amino acid
CC        TOPOLOGY:  linear
CC        MOLECULE TYPE:  protein
CC        ORIGINAL SOURCE:
CC          ORGANISM:  Mus musculus
CC        IMMEDIATE SOURCE:
CC          CLONE:  1.1.1 (murine neuroD2)
SQ   SEQUENCE   382 AA; 41422 MW; 708114 CN;

  Query Match            16.0%;  Score 91;  DB 1;  Length 382;
  Best Local Similarity  29.7%;  Pred. No. 2.14e+00;
  Matches   11;  Conservative   10;  Mismatches  15;  Indels   1;  Gaps   1;

Db     216 FLTEQGADGACRFHGSGGPFAMHPYPYPGSRLAGAQC 252
           :|:   :    | |  : ::  |||  ||: |:::|| 
Qy      33 LLAATETSTIERAHNESPSYIRHPYR-PCGLLSSSQC 68


RESULT    3
ID   5212074-1      STANDARD;      PRT;   322 AA.
XX
AC   xxxxxx
XX
DT
XX
DE  Patent No. 5212074
XX
CC   Patent No. 5212074
CC      APPLICANT: KIEFER, MICHAEL C.;MASIARZ, FRANK R.
CC      TITLE OF INVENTION: GENETIC MATERIAL ENCODING NEW
CC   INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN IGFBP-6
CC      NUMBER OF SEQUENCES: 7
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/07/576,629
CC        FILING DATE: 31-AUG-1990
CC      PRIOR APPLICATION DATA:
CC        APPLICATION NUMBER: 574,614
CC        FILING DATE: 28-AUG-1990
CC   SEQ ID NO:1:
CC        LENGTH: 322
SQ   SEQUENCE   322 AA; 33983 MW; 508193 CN;

  Query Match           14.4%;  Score 82;  DB 5;  Length 322;
  Best Local Similarity  50.0%;  Pred. No. 1.23e+01;
  Matches   11;  Conservative    6;  Mismatches   5;  Indels   0;  Gaps   0;

Db      28 LPPLLLLALLLAASPGGALAR 49
           |||:||:  |||||:  ::: |
Qy      23 LPPFLLPKNLLLAATETSTIER 44


RESULT    4
ID   US-08-690-473-2      STANDARD;      PRT;   1298 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application US/08690473
XX
CC    Sequence 2, Application US/08690473
CC    Patent No. 5676923
CC    GENERAL INFORMATION:
CC      APPLICANT:  Leopardi, Rosario
CC      APPLICANT:  Roizman, Bernard
CC      TITLE OF INVENTION:  HERPES SIMPLEX VIRUS ICP4 AS AN
CC      TITLE OF INVENTION:  INHIBITOR OF APOPTOSIS
CC      NUMBER OF SEQUENCES:  2
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Arnold, White & Durkee
CC        STREET:  P.O. Box 4433
CC        CITY:  Houston
CC        STATE:  Texas
CC        COUNTRY:  USA
CC        ZIP:  77210
```

```
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC DOS/MS DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/690,473
CC         FILING DATE:  26-JUL-1996
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Highlander, Steven L.
CC         REGISTRATION NUMBER:  37,642
CC         REFERENCE/DOCKET NUMBER:  APCD:239
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  512/418-3000
CC         TELEFAX:  512/474-7577
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1298 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:
CC         TOPOLOGY:  linear
SQ   SEQUENCE   1298 AA; 132942 MW; 7107054 CN;

  Query Match           14.4%;  Score 82;  DB 2;  Length 1298;
  Best Local Similarity  28.4%;  Pred. No. 1.23e+01;
  Matches   19;  Conservative   19;  Mismatches  25;  Indels   4;  Gaps   4;

Db   1000 LLSTRDLAFAGAVFFTCLIASAGDRRTIVVNI-VRACDWP-AIMQPAVSPQRAVLACELLP 1057
            || | :: || | :  :| :    : ||:  |   : :::  ::|:  | |:| :| ||:
Qy      7 LLWTTQF-FANARQSTSIPFFLIFFNIILAATFTSTIEFARNESFSYIR-IFYFFCGLLS 64

Db   1058 AVQCAVR 1064
            : ||  |
Qy     65 SSQCLER 71


RESULT    5
ID   US-08-519-547A-6     STANDARD;     PRT;   1130 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 6, Application US/08519547A
XX
CC    Sequence 6, Application US/08519547A
CC    Patent No. 5994082
CC     GENERAL INFORMATION:
CC       APPLICANT:
CC       TITLE OF INVENTION:  Proteins Essential for the Expression of
CC       TITLE OF INVENTION:  Vertebrate MHC Class II Genes, DNA Sequences Encoding Same
CC       TITLE OF INVENTION:  and Pharmaceutical Compositions
CC       NUMBER OF SEQUENCES:  6
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  FISH & NEAVE
CC         STREET:  1251 AVENUE OF THE AMERICAS
CC         CITY:  NEW YORK
CC         STATE:  NEW YORK
CC         COUNTRY:  U.S.A.
CC         ZIP:  10020-1104
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  MS-DOS
CC         SOFTWARE:  WordPerfect 6.1
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/519,547A
CC         FILING DATE:  25-AUG-1995
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  EP94113378.7
CC         FILING DATE:  26-AUG-1994
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  HALEY, JAMES F.
CC         REGISTRATION NUMBER:  27,794
CC         REFERENCE/DOCKET NUMBER:  VOS-11
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  212-596-9000
CC         TELEFAX:  212-596-9090
CC     INFORMATION FOR SEQ ID NO:  6:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1130 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  double
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  DNA (genomic)
CC       HYPOTHETICAL:  NO
CC       ANTI-SENSE:  NO
SQ   SEQUENCE   1130 AA; 123455 MW; 7274599 CN;

  Query Match           13.9%;  Score 79;  DB 2;  Length
  Best Local Similarity  29.3%;  Pred. No. 2.18e+01;
  Matches   12;  Conservative   16;  Mismatches  21;

Db    773 LnAiSAAASVDRNQNVcAFLNe tgr GIERAQQLEEIH 81
          |||:: ::::| :: : :::: :: | | : : 4 ||:
Qy     33 LLAATETSIIERARNEMCYTIGYTR-GLLSSSQCLERLH 71


RESULT    6
ID   US-08-531-525-48      STANDARD;      PRT;   204 AA.
XX
AC   AAAAAA
XX
DT
XX
DE   Sequence 48, Application US/08531525
XX
CC    Sequence 48, Application US/08531525
CC    Patent No. 5840683
CC     GENERAL INFORMATION:
CC       APPLICANT:  Blaxka, Joseph J.
CC       APPLICANT:  Pincus, Matthew R.
CC       APPLICANT:  No, Sandra Yee, John F.
CC       APPLICANT:  Abajian, Henry H.
CC       APPLICANT:  Kondo, Andrew S.
CC       TITLE OF INVENTION:  Peptides Inhibiting the
CC       TITLE OF INVENTION:  of p21 Ras
CC       NUMBER OF SEQUENCES:  52
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Greenlee and Winner, P.C.
CC         STREET:  5370 Manhattan Circle, Suite 201
CC         CITY:  Boulder
CC         STATE:  Colorado
CC         COUNTRY:  US
CC         ZIP:  80303
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC DOS/MS DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/531,525
CC         FILING DATE:  21-SEP-1995
CC         CLASSIFICATION:  530
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Ferber, Joanna M.
CC         REGISTRATION NUMBER:  33,878
CC         REFERENCE/DOCKET NUMBER:  97-94
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (303) 499-8080
CC         TELEFAX:  (303) 499-8089
CC     INFORMATION FOR SEQ ID NO:  48:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  204 amino acids
```

```
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
CC       HYPOTHETICAL:  NO
CC       ORIGINAL SOURCE:
CC         ORGANISM:  Caenorhabditis elegans
SQ   SEQUENCE   204 AA; 23311 MW; 211765 CN;

  Query Match            13.7%;  Score 78;  DB 2;  Length 204;
  Best Local Similarity  25.9%;  Pred. No. 2.63e+01;
  Matches   15;  Conservative   18;  Mismatches  21;  Indels   4;  Gaps   4;

Db    146 WDAKFVEITARESNRVHEVFELLLKETEISRGNLSPTERPNGNSPKRPFKDDGKPCSI 203
          | : |    ||:|  :   | ||:::: ::  : |  ||:: :||   ::   :|||::
Qy      9 WTTQFFA-NAPQSTSLPP-F-LLFKNLLLAATETSTIEPAHNESPSY-IRHPYRPCGL 62


RESULT    7
ID   US-08-718-270A-48      STANDARD;      PRT;   204 AA.
XX
AC   AAAAAA
XX
DT
XX
DE   Sequence 48, Application US/08718270A
XX
CC    Sequence 48, Application US/08718270A
CC    Patent No. 5910478
CC     GENERAL INFORMATION:
CC       APPLICANT:  Hlavka, Joseph J.
CC       APPLICANT:  Pincus, Matthew R.
CC       APPLICANT:  No, 59104781e, John F.
CC       APPLICANT:  Abajian, Henry B.
CC       APPLICANT:  Kende, Andrew S.
CC       TITLE OF INVENTION:  Peptidomimetics Inhibiting
CC       TITLE OF INVENTION:  the Oncogenic Action of P21 Ras
CC       NUMBER OF SEQUENCES:  52
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Greenlee, Winner and Sullivan, P.C.
CC         STREET:  5370 Manhattan Circle, Suite 201
CC         CITY:  Boulder
CC         STATE:  Colorado
CC         COUNTRY:  US
CC         ZIP:  80303
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/718,270A
CC         FILING DATE:  20-SEP-1996
CC         CLASSIFICATION:  514
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 08/531,525
CC         FILING DATE:  21-SEP-1995
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 60/004,041
CC         FILING DATE:  21-SEP-1995
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Ferber, Donna M.
CC         REGISTRATION NUMBER:  33,878
CC         REFERENCE/DOCKET NUMBER:  78-95
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (303) 499-8080
CC         TELEFAX:  (303) 499 8089
CC     INFORMATION FOR SEQ ID NO:  48:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  204 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
CC       HYPOTHETICAL:  NO
CC       ORIGINAL SOURCE:
CC         ORGANISM:  Caenorhabditis elegans
SQ   SEQUENCE   204 AA; 23311 MW; 211765 CN;

  Query Match            13.7%;  Score 78;  DB 2;  Length 204;
  Best Local Similarity  25.9%;  Pred. No. 2.63e+01;
  Matches   15;  Conservative   18;  Mismatches  21;  Indels   4;  Gaps   4;

Db    146 WDAKFVEITARESNRVHEVFELLLKETEISRGNLSPTERPNGNSPKRPFKDDGKPCSI 203
          | : |    ||:|  :   | ||:::: ::  : |  ||:: :||   ::   :|||::
Qy      9 WTTQFFA-NAPQSTSLPP-F-LLFKNLLLAATETSTIEPAHNESPSY-IRHPYRPCGL 62


RESULT    8
ID   US-08-878-283-2      STANDARD;      PRT;   238 AA.
XX
AC   AAAAAA
XX
DT
XX
DE   Sequence 2, Application US/08878283
XX
CC    Sequence 2, Application US/08878283
CC    Patent No. 5852171
CC     GENERAL INFORMATION:
CC       APPLICANT:  Fukudome, Kenji
CC       APPLICANT:  Esmon, Charles T.
CC       TITLE OF INVENTION:  Cloning and Regulation of an Endothelial
CC       TITLE OF INVENTION:  Cell Protein C/Activated Protein C Receptor
CC       NUMBER OF SEQUENCES:  6
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Patrea L. Pabst
CC         STREET:  2800 One Atlantic Center, 1201 West Peachtree
CC         STREET:  Street
CC         CITY:  Atlanta
CC         STATE:  Georgia
CC         COUNTRY:  US
CC         ZIP:  30309-3450
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/878,283
CC         FILING DATE:
CC         CLASSIFICATION:  424
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 08/289,699
CC         FILING DATE:  12-AUG-1994
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Pabst, Patrea L.
CC         REGISTRATION NUMBER:  31,284
CC         REFERENCE/DOCKET NUMBER:  CMEP152
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (404)873-8794
CC         TELEFAX:  (404)873 8795
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  238 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
CC       HYPOTHETICAL:  NO
CC       FEATURE:
CC         NAME/KEY:  misc_feature
CC         LOCATION:  1..365
CC         OTHER INFORMATION:  /note= "Endothelial Cell Protein Receptor
CC         OTHER INFORMATION:  encoded by
CC         OTHER INFORMATION:  nucleotides 1 through 1402 of Sequence ID No. 5852171
CC    Patent No. 5852171
```

```
CC       FEATURE:
CC         NAME/KEY:  Modified-site
CC         LOCATION:  1..15
CC         OTHER INFORMATION:  /note= "Amino acids 1-15 represent
CC         OTHER INFORMATION:  a putative signal sequence."
CC       FEATURE:
CC         NAME/KEY:  Domain
CC         LOCATION:  211..236
CC         OTHER INFORMATION:  /note= "Amino acids 211 236
CC         OTHER INFORMATION:  represent a putative transmembrane domain."
CC       FEATURE:
CC         NAME/KEY:  Active site
CC         LOCATION:  47..174
CC         OTHER INFORMATION:  /note= "Amino acids 47-49, 64 66,
CC         OTHER INFORMATION:  136-138 and 172 174 represent potential
CC         OTHER INFORMATION:  N-glycosylation sites."
CC       FEATURE:
CC         NAME/KEY:  Modified-site
CC         LOCATION:  17..186
CC         OTHER INFORMATION:  /note= "Amino acids 17, 114, 118
CC         OTHER INFORMATION:  and 186 represent extracellular cysteine
CC         OTHER INFORMATION:  residues."
SQ   SEQUENCE   238 AA; 26641 MW; 304069 CN;

  Query Match           13.7%;  Score 78;  DB 2;  Length 238;
  Best Local Similarity  33.3%;  Pred. No. 2.63e-01;
  Matches   13;  Conservative   11;  Mismatches  14;  Indels   1;  Gaps   1;

Db       3 TILLPILLLSGWAFCSQDASDGLQKLHMLQISYFRDPYH 41
           |:| |:||:    : : ::|  ::| |    ||:| ||:
Qy      21 TSLPPFLLFKNLLLAATEIST IEKAHNESPSYIKHPYR 58


RESULT    9
ID   US-09-082-021-2     STANDARD;     PRT:   238 AA.
XX
AC   XXXXXX
XX
DT
XX
DE   Sequence 2, Application US/09082021
XX
CC    Sequence 2, Application US/09082021
CC    Patent No. 6047450
CC     GENERAL INFORMATION:
CC       APPLICANT:  Charles T. Esmon, Deborah J.Stearns Kurosawa and
CC       APPLICANT:  Shinichiro Kurosawa
CC       TITLE OF INVENTION:  DIAGNOSTIC ASSAYS USING SOLUBLE ENDOTHELIAL
CC       TITLE OF INVENTION:  CELL PROTEIN C/ACTIVATED PROTEIN C RECEPTOR
CC       NUMBER OF SEQUENCES:  8
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Patrea L. Pabst
CC         STREET:  2800 One Atlantic Center
CC         STREET:  1201 West Peachtree Street
CC         CITY:  Atlanta
CC         STATE:  Georgia
CC         COUNTRY:  USA
CC         ZIP:  30309-3450
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/09/082,021
CC         FILING DATE:
CC         CLASSIFICATION:
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  08/884,203
CC         FILING DATE:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Pabst, Patrea L.
CC         REGISTRATION NUMBER:  31,284
CC         REFERENCE/DOCKET NUMBER:  MKF168
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (404) 873-8794
CC         TELEFAX:  (404) 873-8795
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  238 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
CC       HYPOTHETICAL:  NO
CC       FEATURE:
CC         NAME/KEY:  misc_feature
CC         LOCATION:  1..469
CC         OTHER INFORMATION:  /note= "Endothelial ce
CC         OTHER INFORMATION:  receptor encoded by nuc
CC    Patent No. 6037450
CC         OTHER INFORMATION:  No. 6,037450 1."
CC       FEATURE:
CC         NAME/KEY:  Modified-site
CC         LOCATION:  1..15
CC         OTHER INFORMATION:  /note= "Amino acids 1-1
CC         OTHER INFORMATION:  a putative signal seque
CC       FEATURE:
CC         NAME/KEY:  Domain
CC         LOCATION:  211..236
CC         OTHER INFORMATION:  /note= "Amino acids 211
CC         OTHER INFORMATION:  represent a putative tr
CC       FEATURE:
CC         NAME/KEY:  Active site
CC         LOCATION:  47..174
CC         OTHER INFORMATION:  /note= "Amino acids 47-
CC         OTHER INFORMATION:  136-138 and 172-174 rep
CC         OTHER INFORMATION:  N-glycosylation sites."
CC       FEATURE:
CC         NAME/KEY:  Active site
CC         LOCATION:  Cys 17
CC         OTHER INFORMATION:  /note immediately proce
CC         OTHER INFORMATION:  cleavage site
CC       FEATURE:
CC         NAME/KEY:  Active site
CC         LOCATION:  64..66
CC         OTHER INFORMATION:  /note peptide inserts
CC         OTHER INFORMATION:  spliced EPCR
CC       FEATURE:
CC         NAME/KEY:  Modified-site
CC         LOCATION:  17..186
CC         OTHER INFORMATION:  /note= "Amino acids 17
CC         OTHER INFORMATION:  and 186 represent extra
SQ   SEQUENCE   238 AA; 26641 MW; 304069 CN;

  Query Match           13.7%;  Score 78;  DB 3;  Len
  Best Local Similarity  33.3%;  Pred. No. 2.63e-01;
  Matches   13;  Conservative   11;  Mismatches  14;

Db       3 TILLPILLLSGWAFCSQDASDGLQKLHMLQISYFRDPYH 41
           |:| |:||:    : : ::|  ::| |    ||:| ||:
Qy      21 TSLPPFLLFKNLLLAATEIST IEKAHNESPSYIKHPYR 58


RESULT   10
ID   US-08-289-699A-2     STANDARD;     PRT:   238 AA
XX
AC   XXXXXX
XX
DT
XX
DE   Sequence 2, Application US/08289699A
XX
CC    Sequence 2, Application US/08289699A
CC    Patent No. 5695993
CC     GENERAL INFORMATION:
CC       APPLICANT:  Fukudome, Kenji
```

```
CC       APPLICANT:  Esmon, Charles T.
CC       TITLE OF INVENTION:  Cloning and Regulation of an Endothelial
CC       TITLE OF INVENTION:  Cell Protein C/Activated Protein C Receptor
CC       NUMBER OF SEQUENCES:  6
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Patrea L. Pabst
CC         STREET:  2800 One Atlantic Center, 1201 West Peachtree
CC         STREET:  Street
CC         CITY:  Atlanta
CC         STATE:  Georgia
CC         COUNTRY:  US
CC         ZIP:  30306-3450
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/289,696A
CC         FILING DATE:  12-AUG-1994
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Pabst, Patrea L.
CC         REGISTRATION NUMBER:  31,284
CC         REFERENCE/DOCKET NUMBER:  OMRF152
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (404)873-8794
CC         TELEFAX:  (404)873-8795
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  238 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
CC       HYPOTHETICAL:  NO
CC       FEATURE:
CC         NAME/KEY:  misc_feature
CC         LOCATION:  1..365
CC         OTHER INFORMATION:  /note= "Endothelial Cell Protein Receptor
CC         OTHER INFORMATION:  encoded by
CC         OTHER INFORMATION:  nucleotides 1 through 1302 of Sequence ID No. 5695993 i."
CC     Patent No. 5695993
CC         FEATURE:
CC         NAME/KEY:  Modified-site
CC         LOCATION:  1..15
CC         OTHER INFORMATION:  /note= "Amino acids 1-15 represent
CC         OTHER INFORMATION:  a putative signal sequence."
CC         FEATURE:
CC         NAME/KEY:  Domain
CC         LOCATION:  211..236
CC         OTHER INFORMATION:  /note= "Amino acids 211-236
CC         OTHER INFORMATION:  represent a putative transmembrane domain."
CC         FEATURE:
CC         NAME/KEY:  Active-site
CC         LOCATION:  47..174
CC         OTHER INFORMATION:  /note= "Amino acids 47-49, 64-66,
CC         OTHER INFORMATION:  136-138 and 172-174 represent potential
CC         OTHER INFORMATION:  N-glycosylation sites."
CC         FEATURE:
CC         NAME/KEY:  Modified-site
CC         LOCATION:  17..186
CC         OTHER INFORMATION:  /note= "Amino acids 17, 114, 118
CC         OTHER INFORMATION:  and 186 represent extracellular cysteine
CC         OTHER INFORMATION:  residues."
SQ   SEQUENCE   238 AA; 26641 MW; 304069 CN;

  Query Match           13.7%;  Score 78;  DB 1;  Length 238;
  Best Local Similarity  33.3%;  Pred. No. 2.63e+01;
  Matches   13;  Conservative   11;  Mismatches   14;  Indels   1,  Gaps   1,

Db       3 TILLPILLLSGWAFCSQDASDCIQRTHMIQISYFRMPYH 41
           |:| ||||:    : : ::|  ::: |    ||:| ||:
Qy      21 TSLPPFLLFKNLLLAATEIST-IERAHNESPSYIRHPYR 58
```

```
RESULT   11
ID   US-08-884-203-2      STANDARD;      PRT;   238 AA.
XX
AC   AAAAAA
XX
DT
XX
DE   Sequence 2, Application US/08884203
XX
CC    Sequence 2, Application US/08884203
CC    Patent No. 5804392
CC     GENERAL INFORMATION:
CC       APPLICANT:  Charles T. Esmon, Deborah J.Stearns-Kurosawa and
CC       APPLICANT:  Shinichiro Kurosawa
CC       TITLE OF INVENTION:  DIAGNOSTIC ASSAYS USING SOLUBLE ENDOTHELIAL
CC       TITLE OF INVENTION:  CELL PROTEIN C/ACTIVATED PROTEIN C RECEPTOR
CC       NUMBER OF SEQUENCES:  8
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Patrea L. Pabst
CC         STREET:  2800 One Atlantic Center
CC         STREET:  1201 West Peachtree Street
CC         CITY:  Atlanta
CC         STATE:  Georgia
CC         COUNTRY:  USA
CC         ZIP:  30309-3450
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/884,203
CC         FILING DATE:
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Pabst, Patrea L.
CC         REGISTRATION NUMBER:  31,284
CC         REFERENCE/DOCKET NUMBER:  OMRF168
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (404) 873-8794
CC         TELEFAX:  (404) 873-8795
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  238 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
CC       HYPOTHETICAL:  NO
CC       FEATURE:
CC         NAME/KEY:  misc_feature
CC         LOCATION:  1..365
CC         OTHER INFORMATION:  /note= "Endothelial Cell Protein
CC         OTHER INFORMATION:  Receptor encoded by nucleotides 1 through 1302 of Sequ
CC     Patent No. 5804392
CC         OTHER INFORMATION:  No. 5804392 1."
CC       FEATURE:
CC         NAME/KEY:  Modified-site
CC         LOCATION:  1..15
CC         OTHER INFORMATION:  /note= "Amino acids 1-15 represent
CC         OTHER INFORMATION:  a putative signal sequence."
CC       FEATURE:
CC         NAME/KEY:  Domain
CC         LOCATION:  211..236
CC         OTHER INFORMATION:  /note= "Amino acids 211-236
CC         OTHER INFORMATION:  represent a putative transmembrane domain."
CC       FEATURE:
CC         NAME/KEY:  Active site
CC         LOCATION:  47..174
CC         OTHER INFORMATION:  /note= "Amino acids 47-49, 64-66,
CC         OTHER INFORMATION:  136-138 and 172-174 represent potential
CC         OTHER INFORMATION:  N-glycosylation sites."
```

```
CC       FEATURE:
CC         NAME/KEY:  Active-site
CC         LOCATION:  Cys 17
CC         OTHER INFORMATION:  /note= immediately, precedes amino acid
CC         OTHER INFORMATION:  cleavage site
CC       FEATURE:
CC         NAME/KEY:  Active-site
CC         LOCATION:  Gly 201
CC         OTHER INFORMATION:  /note-peptide inserts in alternatively
CC         OTHER INFORMATION:  spliced EPCR
CC       FEATURE:
CC         NAME/KEY:  Modified-site
CC         LOCATION:  17..186
CC         OTHER INFORMATION:  /note= "Amino acids 17, 114, 118
CC         OTHER INFORMATION:  and 186 represent extracellular cysteine residues."
SQ   SEQUENCE   238 AA; 26641 MW; 404069 CN;

  Query Match             13.7%;  Score 78;  DB 1;  Length 238;
  Best Local Similarity  33.3%;  Pred. No. 2.6e-01;
  Matches   13;  Conservative   11;  Mismatches  14;  Indels   1;  Gaps   1;

Db       8 TTLLPILLLSGWAFCSQDASDGLQRLHMLQISYFKDPYH 41
           |:| |:||:    : : :::  ::: |    ||:| ||:
Qy      21 ISLPPFLLFKNLLLAAFETST-IERAHNESPSYIKHPYR 58


RESULT   12
ID   US-08-241-387-5      STANDARD;      PRT;   141 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 5, Application US/08241387
XX
CC    Sequence 5, Application US/08241387
CC    Patent No. 5589570
CC     GENERAL INFORMATION:
CC       APPLICANT:  Tamura, Richard N.
CC       APPLICANT:  Quaranta, Vito
CC       TITLE OF INVENTION:  INTEGRIN ALPHA SUBUNIT CYTOPLASMIC
CC       TITLE OF INVENTION:  DOMAIN POLYPEPTIDES, ANTIBODIES AND METHODS
CC       NUMBER OF SEQUENCES:  16
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  The Scripps Research Institute
CC         STREET:  10666 No. 55895701h Torrey Pines Road, TPC-8
CC         CITY:  La Jolla
CC         STATE:  CA
CC         COUNTRY:  US
CC         ZIP:  92037
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  Patentin Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  08/241,387
CC         FILING DATE:  10-MAY-1994
CC         CLASSIFICATION:  530
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US08/07/695,564
CC         FILING DATE:  03-MAY-1004
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Fitting, Thomas
CC         REGISTRATION NUMBER:  34,163
CC         REFERENCE/DOCKET NUMBER:  TSR1241.0D1
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  619-554-2937
CC         TELEFAX:  619-554-6312
CC     INFORMATION FOR SEQ ID NO:  5:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  141 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  unknown
CC       MOLECULE TYPE:  protein
CC       HYPOTHETICAL:  YES
CC       ANTI SENSE:  No
CC       FRAGMENT TYPE:  internal
CC       FEATURE:
CC         NAME/KEY:  Region
CC         LOCATION:  1..141
CC         OTHER INFORMATION:  /note= "The 141 amino ac
CC         OTHER INFORMATION:  predicted from the nucl
CC         OTHER INFORMATION:  results from amplificat
CC       FEATURE:
CC         NAME/KEY:  Domain
CC         LOCATION:  88..113
CC         OTHER INFORMATION:  /note= "The putative tr
CC         OTHER INFORMATION:  domain."
CC       FEATURE:
CC         NAME/KEY:  Region
CC         LOCATION:  1..126
CC         OTHER INFORMATION:  /note= "SEQ ID NO:5 is
CC         OTHER INFORMATION:  SEQ ID NO:7 at amino ac
CC         OTHER INFORMATION:  the two sequences differ
SQ   SEQUENCE   141 AA; 16105 MW; 3669520 CN;

  Query Match             16.4%;  Score 76;  DB 1;  Len
  Best Local Similarity  43.5%;  Pred. No. 3.8e-01;
  Matches   10;  Conservative    3;  Mismatches   9;

Db     109 VFSLLWKCGFFKRSRYDDSIPKY 131
           || |||   ||  :|   |:. :
Qy       4 VFSLLWTIQFFANARQSISLPFF 26


RESULT   13
ID   US-07-695-564-5      STANDARD;      PRT;   141 AA
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 5, Application US/07695564
XX
CC    Sequence 5, Application US/ 7695564
CC    Patent No. 5310874
CC     GENERAL INFORMATION:
CC       APPLICANT:  Tamura, Richard N.
CC       APPLICANT:  Quaranta, Vito
CC       TITLE OF INVENTION:  INTEGRIN ALPHA SUBUNIT CY
CC       TITLE OF INVENTION:  DOMAIN POLYPEPTIDES, ANTI
CC       NUMBER OF SEQUENCES:  16
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Thomas Fitting
CC         STREET:  11300 Sorrento Valley Road, Suite .
CC         CITY:  San Diego
CC         STATE:  California
CC         COUNTRY:  United States
CC         ZIP:  92121
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  Patentin Release #1.0, Versi: #
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/07/695,564
CC         FILING DATE:  1991
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:
CC         FILING DATE:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Fitting, Thomas
CC         REGISTRATION NUMBER:  34,163
CC         REFERENCE/DOCKET NUMBER:  SCRO377P
```

```
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  619 546 1555
CC     INFORMATION FOR SEQ ID NO:  5:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  141 amino acids
CC         TYPE:  AMINO ACID
CC         TOPOLOGY:  unknown
CC       MOLECULE TYPE:  protein
CC       HYPOTHETICAL:  YES
CC       ANTI-SENSE:  NO
CC       FRAGMENT TYPE:  internal
CC       FEATURE:
CC         NAME/KEY:  Region
CC         LOCATION:  1..141
CC         OTHER INFORMATION:  /note= "The 141 amino acid sequence
CC         OTHER INFORMATION:  predicted from the nucleic acid product which
CC         OTHER INFORMATION:  results from amplification of the mouse ALPHA 6b
CC         OTHER INFORMATION:  cDNA with primers 1157/1156."
CC       FEATURE:
CC         NAME/KEY:  Domain
CC         LOCATION:  88..113
CC         OTHER INFORMATION:  /note= "The putative transmembrane
CC         OTHER INFORMATION:  domain."
CC       FEATURE:
CC         NAME/KEY:  Region
CC         LOCATION:  1..120
CC         OTHER INFORMATION:  /note= "SEQ ID NO:5 is identical to
CC         OTHER INFORMATION:  SEQ ID NO:7 at amino acid position 1 through 120;
CC         OTHER INFORMATION:  the two sequences diverge at amino acid 121."
SQ   SEQUENCE   141 AA; 16105 MW; 103520 CN;

  Query Match           13.4%;  Score 76;  DB 1;  Length 141;
  Best Local Similarity  43.5%;  Pred. No. 3.83e+01;
  Matches   10;  Conservative    3;  Mismatches   9;  Indels   1;  Gaps   1;

Db    109 VF-LLWKCGFFKRSRYDDSIPRY 130
          || |||   ||  :|   |:| :
Qy      4 VFSLIWTTQFFANARQSTSLPPF 26


RESULT   14
ID   US-08-460-907B-2     STANDARD;     PRT;   1531 AA.
XX
AC   XXXXXX
XX
DT
XX
DE   Sequence 2, Application US/08460907B
XX
CC    Sequence 2, Application US/08460907B
CC    Patent No. 5891724
CC     GENERAL INFORMATION:
CC       APPLICANT:  Deeley, Roger G.
CC       APPLICANT:  Cole, Susan P.C.
CC       TITLE OF INVENTION:  METHODS FOR CONFERRING MULTIDRUG
CC       TITLE OF INVENTION:  RESISTANCE ON A CELL
CC       NUMBER OF SEQUENCES:  9
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  PARTEQ RESEARCH & DEVELOPMENT INNOVATIONS
CC         STREET:  Queen's University at Kingston
CC         CITY:  Kingston
CC         STATE:  Ontario
CC         COUNTRY:  CANADA
CC         ZIP:  K7L 3N6
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  ASCII text
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/460,907B
CC         FILING DATE:  05-JUN-1995
CC         CLASSIFICATION:  424
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  07/966,923
CC         FILING DATE:  27-OCT-1992
CC         CLASSIFICATION:  424
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  08/029,340
CC         FILING DATE:  8-MAR-1993
CC         CLASSIFICATION:  424
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  08/141,893
CC         FILING DATE:  26-OCT-1993
CC         CLASSIFICATION:  424
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  08/407,207
CC         FILING DATE:  20-MAR-1995
CC         CLASSIFICATION:  424
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Steeg, Carol Miernicki
CC         REGISTRATION NUMBER:  39,539
CC         REFERENCE/DOCKET NUMBER:  Q1551
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (613) 545 2342
CC         TELEFAX:  (613) 545-6853
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1531 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   1531 AA; 171671 MW; 12353123 CN;

  Query Match           13.4%;  Score 76;  DB 2;  Length 1531;
  Best Local Similarity  27.3%;  Pred. No. 3.83e+01;
  Matches   15;  Conservative   14;  Mismatches  24;  Indels   2;  Gaps   2;

Db   1435 RQLVCLAKALLRKTKILVLDEATAAVDLETDDLIQSTIRTQFEDCTVITIAHRLN 1489
          ||   |:  || |  :|   | |::::   ::   | ||  :  | :|: :: |:
Qy     18 RQSTSLPFFLLFKNLLLAAIA-ISIIERAANE-SPSYIKHPYRPCGLLSSSQCLE 70


RESULT   15
ID   US 08 461 384B 2     STANDARD;     PRT;   1531 AA.
XX
AC   XXXXXX
XX
DT
XX
DE   Sequence 2, Application US/08461384B
XX
CC    Sequence 2, Application US/08461384B
CC    Patent No. 6025473
CC     GENERAL INFORMATION:
CC       APPLICANT:  Cole, Susan P.C.
CC       APPLICANT:  Deeley, Roger G.
CC       TITLE OF INVENTION:  MULTIDRUG RESISTANCE PROTEINS
CC       NUMBER OF SEQUENCES:  10
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  PARTEQ RESEARCH & DEVELOPMENT INNOVATIONS
CC         STREET:  Queen's University at Kingston
CC         CITY:  Kingston
CC         STATE:  Ontario
CC         COUNTRY:  CANADA
CC         ZIP:  K7L 3N6
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  ASCII text
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/461,384B
CC         FILING DATE:  05-JUN-95
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  07/966,923
```

CC FILING DATE: 27-OCT-1992
CC APPLICATION NUMBER: 08/029,340
CC FILING DATE: 8-MAR-1993
CC APPLICATION NUMBER: 08/141,893
CC FILING DATE: 26-OCT-1993
CC APPLICATION NUMBER: 08/407,207
CC FILING DATE: 20-MAR-1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Steeg, Carol Miernicki
CC REGISTRATION NUMBER: 39,539
CC REFERENCE/DOCKET NUMBER: 01547
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (613) 545-2342
CC TELEFAX: (613) 545-6853
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1531 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 1531 AA; 171671 MW; 12353123 CN;

Query Match 13.4%; Score 76; DB 3; Length 1531;
Best Local Similarity 27.3%; Pred. No. 3.83e+01;
Matches 15; Conservative 14; Mismatches 24; Indels 2; Gaps 2;

Db 1435 RQIVCIARAIIRKTKIIVIDFATAAVDLFTDDLIQSTIRTQFEDCIVLTIAHRLN 1489
Qy 18 RQSTSLPPFLLFKNLLLAATE-TSTIERAHNE-SPSYIRHPYRPCGLLSSSQCLE 70

Search completed: Fri Aug 25 17:51:30 2000
Job time : 20 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search using Smith-Waterman algorithm

Run on: Fri Aug 25 17:48:14 2000; MasPar time 6.61 Seconds
549.711 Million cell updates/sec

Tabular output not generated.

Title: >US-09-257-585-3
Description: (1-77) from US09257585.pep
Perfect Score: 568
Sequence: 1 MGPVFSLLWTIQFFANARQS.....KPCGLLSSSQCLERLHHPTL 77

Scoring table: PAM 150
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir64
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 38.246; Variance 70.385; scale 0.543

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	103	18.1	971	1	A36376	replication licensing	3.21e-03
2	92	16.2	1623	2	I01369	multidrug-resistance	1.44e-01
3	91	16.0	381	2	G02668	neurogenic basic-heli	2.01e-01
4	91	16.0	383	2	JC4688	neuro D-related facto	2.01e-01
5	90	15.8	1861	2	T13845	microtubule-associate	2.80e-01
6	89	15.7	140	2	T16574	hypothetical protein	3.90e-01
7	88	15.5	533	2	T15116	hypothetical protein	5.41e-01
8	87	15.3	221	2	G70381	RNase III - Aquifex a	7.48e-01
9	86	15.1	219	2	D70411	hypothetical protein	1.03e+00
10	84	14.8	389	2	T03411	protein RFL - rice	1.95e+00
11	83	14.6	157	2	S49622	ertK protein - Rhodob	2.67e+00
12	83	14.6	158	2	A57448	tryptophan-rich senso	2.67e+00
13	83	14.6	241	2	S77548	lumQ protein - Synech	2.67e+00
14	83	14.6	278	2	S49623	ertC protein - Rhodob	2.67e+00
15	82	14.4	238	2	A71216	hypothetical protein	3.65e+00
16	82	14.4	240	2	A39842	insulin-like growth f	3.65e+00
17	82	14.4	375	2	B71842	carbamoyl phosphate s	3.65e+00
18	82	14.4	1298	1	EDBE75	immediate-early prote	3.65e+00
19	81	14.3	207	2	S40747	hypothetical protein	4.98e+00
20	81	14.3	256	2	D69042	cobalt transport prot	4.98e+00
21	81	14.3	394	2	S75004	hypothetical protein	4.98e+00
22	80	14.1	515	2	G72097	hypothetical protein	6.76e+00
23	80	14.1	541	2	A49801	3-hydroxyacylphosphoe	6.76e+00
24	80	14.1	637	2	S61509	NADH dehydr	6.76e+00
25	80	14.1	638	2	I31891	NADH dehydr	6.76e+00
26	80	14.1	644	1	A40212	uromodulin	6.76e+00
27	80	14.1	644	2	JN0644	lamin-associ	6.76e+00
28	80	14.1	650	2	S47291	probable no	6.76e+00
29	80	14.1	699	1	HEHU65	DNA recomb	6.76e+00
30	80	14.1	780	2	I02939	chloride ch	6.76e+00
31	79	13.9	156	1	R5HU67	ribosomal	8.83e+00
32	79	13.9	332	2	B69412	nitrate re	8.83e+00
33	79	13.9	444	2	S54409	membrane p	8.83e+00
34	79	13.9	565	1	S09915	ribosomal	8.83e+00
35	79	13.9	1140	2	A48843	MHC class	8.83e+00
36	78	13.7	238	2	A55869	endothelia	1.16e-01
37	78	13.7	373	2	B71523	probable ab	1.16e-01
38	78	13.7	413	2	I31534	hypothetica	1.16e-01
39	78	13.7	541	2	B54429	2 hydroxya	1.16e-01
40	78	13.7	658	1	F60JSA	gag polypr	1.16e-01
41	78	13.7	699	2	I32719	NADH dehydr	1.16e-01
42	77	13.6	338	2	S16618	opacity pr	1.54e-01
43	77	13.6	446	2	I37695	farnesyl-t	1.54e-01
44	77	13.6	517	2	S31042	cytochrome	1.54e-01
45	77	13.6	822	2	I01622	hypothetic	1.54e-01

ALIGNMENTS

RESULT 1
ENTRY A36376 #type complete
TITLE replication licensing factor MCM3 - ye
cerevisiae)
ALTERNATE_NAMES minichromosome maintenance protein MCM
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 13-Aug-1990 #sequence revision 13-Aug
20-Aug-1990
ACCESSIONS A36376; S50512; S30851
REFERENCE A36376
#authors Gibson, S.I.; Surosky, R.T.; Tye, B.K
#journal Mol. Cell. Biol. (1990) 10:5707-5720
#title The phenotype of the minichromosome ma
is characteristic of mutants defect
#cross-references MUID:91042499
#accession A36376
##molecule_type DNA
##residues 1-971 ##label GIB
##cross-references GB:X54543; NID:g39914; PID:g
REFERENCE S50491
#authors Dietrich, F.S.
#submission submitted to the EMBL Data Library, J
#description The sequence of S. cerevisiae cosmids
#accession S50512
##molecule_type DNA
##residues 1-971 ##label JE
##cross-references EMBL:Z16779; NID:g603626; PID
PID:g603647; MIPS:YEL042w
REFERENCE S30812
#authors Mulligan, J.T.; Dietrich, F.S.; Hennes
Komp, C.; Wei, Y.; Taylor, P.; Naka
Davis, R.W.
#submission submitted to the EMBL Data Library, F
#accession S30851
##molecule_type DNA
##residues 1-954 ##label MUL
##cross-references EMBL:L10430
COMMENT The complex of six MCM proteins is one of s
must be bound at the origins of repli
DNA replication. As replication proceeds,
phosphorylated and dissociate from the ch
GENETICS
#gene SGD:MCM3
##cross-references SGD:S0005075; MIPS:YEL042w
#map_position 5L
COMPLEX The predominant form is a heterohexame
MCM1, MCM2, and MCM3, sometimes cal

```
                  of replication licensing factor.
FUNCTION
   #description   part of the replication licensing system that permits DNA
                  replication to occur once and only once within a single
                  cell division cycle
CLASSIFICATION    #superfamily yeast replication licensing factor MCM3; MCM
                  homology
KEYWORDS          cell cycle control; DNA replication initiation; nucleus;
                  phosphoprotein
FEATURE
   168-742             #domain MCM homology #label MCM
SUMMARY           #length 971  #molecular-weight 107516  #checksum 6721

  Query Match             18.1%;  Score 103;  DB 1;  Length 971;
  Best Local Similarity   41.7%;  Pred. No. 3.2le-03;
  Matches    19;  Conservative    14;  Mismatches   25;  Indels    2,  Gaps    2,

Db     528 DPHQNIALPDSLLSRFDLLFVVTDDINEIRDRSISEHVLRTHRYLPPGYLEGEPVRERLN 587
           :::|: :|| ||  : :||:  |:     | :: |   :| | | | | | :    |||:
Qy      16 NARQSTSLPPFLLFK-NLLLAATFTSTIEPAHNESPSYIR-HDYPPCCLLSSSQCLERLH 73


RESULT    2
ENTRY             T01369      #type complete
TITLE             multidrug-resistance protein homolog T29F13.13 - Arabidopsis
                  thaliana
ORGANISM          #formal_name Arabidopsis thaliana #common_name mouse ear
                  cress
DATE              12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
                  30-Apr-1999
ACCESSIONS        T01369
REFERENCE         214177
   #authors       Rounsley, S.D.; Liu, X.; Ketchum, K.A.; Crosby, M.L.;
                  Brandon, R.C.; Sykes, S.M.; Kaul, S.; Mason, T.M.;
                  Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter,
                  J.C.
   #submission    submitted to the EMBL Data Library, May 1998
   #description   Arabidopsis thaliana chromosome II BAC T29F13 genomic
                  sequence.
   #accession     T01369
      ##status          translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues       1-1623 ##label ROU
      ##cross-references EMBL:AC002096; NID:g2172469; PID:g3132179
      ##experimental_source cultivar Columbia
GENETICS
   #map_position 2
   #introns       110/3; 170/3; 183/3; 258/3; 274/2; ...; 450/3; 577/3; 633/3; 656/3; ...; 693/3; 726/3; 777/3; 918/2; 979/3; 1078/3; 1371/3; 1456/3; 1592/1
   #note          T29F13.13
SUMMARY           #length 1623  #molecular-weight 182128  #checksum 624

  Query Match             16.2%;  Score 92;  DB 2;  Length 1623;
  Best Local Similarity   26.7%;  Pred. No. 1.44e-01;
  Matches    16;  Conservative    16;  Mismatches   26;  Indels    2,  Gaps    2,

Db    1379 FSVGQRQLLSLSRALLRRSKILVIDEATAAVDVRTDALIQKTIREEFKSCTMLIIAHRLN 1438
           |    ||  ||:  || :: :|   | |:::: |:       ||  :::| :|  :: |:
Qy      13 FFANARQSTSLPPFLLFKNLLLAAIE-TSTIE-PAHNESPSYIRHDYPPCCLLSSSQCLE 70


RESULT    3
ENTRY             G02668      #type complete
TITLE             neurogenic basic-helix-loop-helix (bHLH) protein - human
ORGANISM          #formal_name Homo sapiens #common_name man
DATE              21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change
                  17-Jul-1998
ACCESSIONS        G02668
REFERENCE         H01573
   #authors       Tapscott, S.J.; Lamimi, R.L.; McCormick, P.M.
   #submission    submitted to the EMBL Data Library, May 1996
--4      3

   #accession     G02668
      ##status          preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues       1-381 ##label TAP
      ##cross-references EMBL:U58681; NID:g1477748; PID:g1477749
GENETICS
   #gene          NeuroD2
SUMMARY           #length 381  #molecular-weight 41300  #checksum 3349

  Query Match             16.0%;  Score 91;  DB 2;  Length 381;
  Best Local Similarity   29.7%;  Pred. No. 2.01e-01;
  Matches    11;  Conservative    10;  Mismatches   15;  Indels    1;  Gaps    1;

Db     215 FLTEQCADGAGPFHGSGGPFAMHPYPYPCSPLAGAQC 251
           :|:   :    | |  : ::  |||  ||: |::::|
Qy      53 LLAATETSTIERAHNESPSYIRHFYR-PCGLLSSSQC 68


RESULT    4
ENTRY             JC4688      #type complete
TITLE             neuro D-related factor - mouse
ORGANISM          #formal_name Mus musculus #common_name house mouse
DATE              10-May-1996 #sequence_revision 19-Jul-1996 #text_change
                  17-Mar-1999
ACCESSIONS        JC4688
REFERENCE         JC4688
   #authors       Yasunami, M.; Suzuki, K.; Maruyama, H.; Kawakami, H.; Nagai,
                  Y.; Hagiwara, M.; Ohkubo, H.
   #journal       Biochem. Biophys. Res. Commun. (1996) 220:754-758
   #title         Molecular cloning and characterization of a cDNA encoding a
                  novel basic helix-loop-helix protein structurally related
                  to Neuro D/BHF1.
   #cross-references MUID:96183021
   #accession     JC4688
      ##molecule_type mRNA
      ##residues       1-383 ##label YAS
      ##cross-references DDBJ:D83507; NID:g1304163; PID:d1012599; PID:g1304164
      ##experimental_source embryo
COMMENT           This factor plays distinct roles in neural development and
                  plasicity as a novel member of the NeuroD family.
FEATURE
   46-48               #region proline-rich\
   83-92               #region glutamic acid-rich\
   213-297             #region alanine/glycine-rich
SUMMARY           #length 383  #molecular-weight 41493  #checksum 389

  Query Match             16.0%;  Score 91;  DB 2;  Length 383;
  Best Local Similarity   29.7%;  Pred. No. 2.01e-01;
  Matches    11;  Conservative    10;  Mismatches   15;  Indels    1;  Gaps    1;

Db     215 FLTEQGADGAGPFHGSGGPFAMHPYPYPCSPLAGAQC 252
           :|:   :    | |  : ::  |||  ||: |::::|
Qy      53 LLAATETSTIERAHNESPSYIRHFYR-PCGLLSSSQC 68


RESULT    5
ENTRY             T13845      #type complete
TITLE             microtubule-associated protein - fruit fly (Drosophila
                  melanogaster)
ORGANISM          #formal_name Drosophila melanogaster
DATE              20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change
                  20-Sep-1999
ACCESSIONS        T13845
REFERENCE         217792
   #authors       Saunders, R.D.; Avides, M.C.; Howard, T.; Gonzalez, C.;
                  Glover, D.M.
   #journal       J. Cell Biol. (1997) 137:881-890
   #title         The Drosophila gene abnormal spindle encodes a
                  microtubule-associated protein that associates with the
                  polar regions of the mitotic spindle.
   #accession     T13845
      ##status          preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
```

```
      ##residues       1-1861 ##label SAU
      ##cross-references EMBL:U95171; NID:g1930121; PID:g1930122;
                        PIDN:AAB51540.1
GENETICS
   #gene           asp
FUNCTION
   #description  is required for the normal function of the mitotic spindle
SUMMARY           #length 1861  #molecular-weight 219558  #checksum 5769

  Query Match             15.8%;  Score 90;  DB 2;  Length 1861;
  Best Local Similarity   44.6%;  Pred. No. 2.80e-01;
  Matches    18;  Conservative    15;  Mismatches   16;  Indels    3;  Gaps    2;

Db    1003 LFAKKLCAAAETARQQRSQKQQAAASYIQMQWRIYQL-GRIQRHEPLRQRIL 1053
           || : ||||:|| :: :::  ::|||:  |   | :  |  | |::  |
Qy      28 LFKNLLLAATETSTIERAHNE-SPSYIRHPYRPCGLLSSSQCLERLHHPIL 77


RESULT    6
ENTRY             I16574      #type complete
TITLE             hypothetical protein K05F1.10 - Caenorhabditis elegans
ORGANISM          #formal_name Caenorhabditis elegans
DATE              20-Sep-1999 #sequence_revision 20-Sep 1999 #text_change
                    20-Sep-1999
ACCESSIONS        I16574
REFERENCE         Z18537
   #authors       Wohldmann, P.
   #submission    submitted to the EMBL Data Library, June 1995
   #description   The sequence of C. elegans cosmid K05F1.
   #accession     I16574
      ##status          preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues       1-140 ##label WOH
      ##cross-references EMBL:U29477; NID:g868173; PID:g868180;
                        PIDN:AAA68717.1; CESP:K05F1.10
      ##experimental_source strain Bristol N2
GENETICS
   #gene           CESP:K05F1.10
   #introns        31/3, 77/2
SUMMARY           #length 140  #molecular-weight 16371  #checksum 8222

  Query Match             15.7%;  Score 89;  DB 2;  Length 140;
  Best Local Similarity   47.8%;  Pred. No. 3.90e-01;
  Matches    11;  Conservative    7;  Mismatches    4;  Indels    1;  Gaps    1;

Db      66 RHCDRTCENLFSPPHCLNHLHHA 88
           ||  | | |  |:|::::||::|||:
Qy      54 RHPYRPC-GLLSSSQCLERLHHP 75


RESULT    7
ENTRY             T15116      #type complete
TITLE             hypothetical protein T23C6.3 - Caenorhabditis elegans
ORGANISM          #formal_name Caenorhabditis elegans
DATE              20-Sep-1999 #sequence_revision 20 Sep 1999 #text_change
                    20-Sep-1999
ACCESSIONS        T15116
REFERENCE         Z18295
   #authors       Du, Z.
   #submission    submitted to the EMBL Data Library, April 1997
   #description   The sequence of C. elegans cosmid T23C6.
   #accession     T15116
      ##status          preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues       1-533 ##label DUZ
      ##cross-references EMBL:AF000191; NID:g1946976; PID:g1946980;
                        PIDN:AAB52884.1; GSPDB:GN00028; CESP:T23C6.3
      ##experimental_source strain Bristol N2; clone T23C6
GENETICS
   #gene           CESP:T23C6.3
   #map_position 10
   #introns        28/3, 257/3, 454/3
SUMMARY           #length 533  #molecular-weight 61479  #checksum 10

  Query Match             15.5%;  Score 88;  DB 2;  Length
  Best Local Similarity   [illegible]%;  Pred. No. 5.41e-01;
  Matches    10;  Conservative    [illegible];  Mismatches   14;  [illegible]

Db    472 PIPTIPSTSTMPTPMPGSLNVTIAA I SVAPIDSLIPSPIP 515
          |::   :::::|| : : ::: : : :  :  : :::::
Qy     13 FFANARQSTSLPTPLLENIAAIIICTIEKAHNESPSYIP 54


RESULT    8
ENTRY             G70381      #type complete
TITLE             RNase III    Aquifex aeolicus
ORGANISM          #formal_name Aquifex aeolicus
DATE              08-May-1998 #sequence_revision 08-May
                    20-Sep-1999
ACCESSIONS        G70381
REFERENCE         A70300
   #authors       Deckert, G.; Warren, P.V.; Gaasterland
                    Lenox, A.L.; Graham, D.E.; Overbeek
                    Keller, M.; Aujay, M.; Huber, R.; F
                    J.M.; Olsen, G.J.; Swanson, R.V.
   #journal       Nature (1998) 392:353-358
   #title         The complete genome of the hyperthermo
                    Aquifex aeolicus.
   #cross-references MUID:98196
   #accession     G70381
      ##status          preliminary; nucleic acid sequen
                         translation not shown
      ##molecule_type DNA
      ##residues       1-221 ##label AQF
      ##cross-references GB:AE000[illegible]; NID:g2983460; P
      ##experimental_source strain VF5
GENETICS
   #gene           rnc
CLASSIFICATION    #superfamily ribonuclease III; double
                    repeat homology
SUMMARY           #length 221  #molecular-weight 25[illegible]

  Query Match             15.7%;  Score 87;  DB 2;  Len
  Best Local Similarity   43.5%;  Pred. No. 7.48e-01;
  Matches    12;  Conservative    7;  Mismatches   10;  [illegible]

Db    105 SSPVTEAIWAAVYLTSGAVSLIRKWFDLSAIEDRV
          :| ||  ||:: :: :: ::  :      ||| :|||  : :
Qy      1 MGPVFSLLWTTQFFANARQSI SLPTPLLEFKNLLLAAIEISII

RESULT    9
ENTRY             D70411      #type complete
TITLE             hypothetical protein aq_1287   Aquifex
ORGANISM          #formal_name Aquifex aeolicus
DATE              08-May-1998 #sequence_revision 08-May
                    08-May-1998
ACCESSIONS        D70411
REFERENCE         A70300
   #authors       Deckert, G.; Warren, P.V.; Gaasterland
                    Lenox, A.L.; Graham, D.E.; Overbeek
                    Keller, M.; Aujay, M.; Huber, R.; F
                    J.M.; Olsen, G.J.; Swanson, R.V.
   #journal       Nature (1998) 392:353-358
   #title         The complete genome of the hyperthermo
                    Aquifex aeolicus.
   #cross-references MUID:98196
   #accession     D70411
      ##status          preliminary; nucleic acid sequen
                         translation not shown
      ##molecule_type DNA
      ##residues       1-219 ##label AQF
      ##cross-references GB:AE000732; NID:g2984704; PI
      ##experimental_source strain VF5
GENETICS
   #gene           aq_1287
```

```
SUMMARY          #length 219  #molecular-weight 25050  #checksum 7458

  Query Match           15.1%;  Score 86;  DB 2;  Length 219;
  Best Local Similarity  43.8%;  Pred. No. 1.03e+00;
  Matches    14;  Conservative    6;  Mismatches  10;  Indels    2;  Gaps    2;

Db      13 VFGLFLFSLIFFGLLSLAT-FPKFLLFDRLLI 43
           ||:| | :  ||:    :| :| |||| ||:
Qy       4 VFSL-LWTTQFFANARQSTSLPPFLLFKNLLL 34


RESULT   10
ENTRY            I03411      #type complete
TITLE            protein RFL - rice
ORGANISM         #formal_name Oryza sativa #common_name rice
DATE             24-Mar-1999 #sequence_revision 24-Mar 1999 #text_change
                   24-Mar-1999
ACCESSIONS       I03411
REFERENCE        Z14938
   #authors      Kyozuka, J.; Konishi, S.; Nemoto, K.; Izawa, T.; Shimamoto,
                   K.
   #submission   submitted to the EMBL Data Library, July 1997
   #description  Downregulation of RFL, rice FLO/LFY homolog, accompanied with
                   panicle branching.
   #accession    I03411
      ##status        preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues      1-389 ##label KYO
      ##cross-references EMBL:AB005620; NID:d1111216; PID:d1022397
      ##experimental_source cv. Toride
GENETICS
   #note         RFL
FUNCTION
   #description  might be involved in panicle branching
SUMMARY          #length 389  #molecular-weight 42530  #checksum 4253

  Query Match           14.8%;  Score 84;  DB 2;  Length 389;
  Best Local Similarity  31.7%;  Pred. No. 1.95e+00;
  Matches    13;  Conservative    9;  Mismatches  18;  Indels    1;  Gaps    1;

Db     225 FVVTEPQIEVARAKKNGLDYLPHLYEQCRLFLLQVQSMAKLH 265
           :  || : : || :::  |: | |  | | |    |: :||
Qy      34 LAATETSTIFKAHNESPSYIRHPYRPCGL-LSSSQCLEMLH 73


RESULT   11
ENTRY            S49622      #type complete
TITLE            crtK protein - Rhodobacter sphaeroides
ORGANISM         #formal_name Rhodobacter sphaeroides
DATE             05-Mar-1995 #sequence_revision 12-May-1995 #text_change
                   04-Sep-1998
ACCESSIONS       S49622
REFERENCE        S49619
   #authors      Lang, H.P.; Coqdell, R.J.; Takaichi, S.; Hunter, C.N.
   #submission   submitted to the EMBL Data Library, November 1994
   #description  The complete DNA sequence, specific INS insertion map and
                   gene assignment of the carotenoid biosynthesis genes of
                   Rhodobacter sphaeroides.
   #accession    S49622
      ##status        preliminary
      ##molecule_type DNA
      ##residues      1-157 ##label LAN
      ##cross-references EMBL:X82458; NID:q575405; PID:q575409
GENETICS
   #gene         crtK
CLASSIFICATION   #superfamily tryptophan-rich sensory protein
SUMMARY          #length 157  #molecular-weight 17854  #checksum 6363

  Query Match           14.6%;  Score 83;  DB 2;  Length 157;
  Best Local Similarity  39.4%;  Pred. No. 2.67e+00;
  Matches    13;  Conservative   10;  Mismatches   9;  Indels    1;  Gaps    1;

Db      82 FNTLWTPVFFGMKRMATALAVVMV-MWLFVAAT 114
           |: ||| ||:  | :|:|:  :: : |::|||
Qy       5 FSLLWTTQFFANARQSTSLPPFLLFKNLLLAAT 37


RESULT   12
ENTRY            A57438      #type complete
TITLE            tryptophan rich sensory protein  Rhodobacter sphaeroides
                   (strain 2.4.1)
ORGANISM         #formal_name Rhodobacter sphaeroides
DATE             26-Jan-1996 #sequence_revision 26-Jan 1996 #text_change
                   04-Sep-1998
ACCESSIONS       A57438
REFERENCE        A57438
   #authors      Yeliseev, A.A.; Kaplan, S.
   #journal      J. Biol. Chem. (1995) 270.21167 21175
   #title        A sensory transducer homologous to the mammalian
                   peripheral type benzodiazepine receptor regulates
                   photosynthetic membrane complex formation in Rhodobacter
                   sphaeroides 2.4.1.
   #cross-references MUID:95403350
   #accession    A57438
      ##status        preliminary; not compared with conceptual translation
      ##molecule_type DNA
      ##residues      1-158 ##label YEL
GENETICS
   #gene         tspO
CLASSIFICATION   #superfamily tryptophan-rich sensory protein
SUMMARY          #length 158  #molecular-weight 17976  #checksum 9469

  Query Match           14.6%;  Score 83;  DB 2;  Length 158;
  Best Local Similarity  39.4%;  Pred. No. 2.67e+00;
  Matches    13;  Conservative   10;  Mismatches   9;  Indels    1;  Gaps    1;

Db      83 FNTLWTPVFFGMKRMATALAVVMV-MWLFVAAT 114
           |: ||| ||:  | :|:|:  :: : |::|||
Qy       5 FSLLWTTQFFANARQSTSLPPFLLFKNLLLAAT 37


RESULT   13
ENTRY            S77548      #type complete
TITLE            hypo protein - Synechocystis sp. (strain PCC 6803)
ALTERNATE_NAMES  protein slr1213
ORGANISM         #formal_name Synechocystis sp.
   #variety      PCC 6803
DATE             25 Apr 1997 #sequence_revision 25 Apr 1997 #text_change
                   05-Dec-1998
ACCESSIONS       S77548
REFERENCE        S74822
   #authors      Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
                   Nakamura, Y., Miyajima, N., Hirosawa, M.; Sugiura, M.;
                   Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
                   Muraki, A., Nakazaki, N., Naruo, K., Okumura, S.; Shimpo,
                   S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
                   Yasuda, M.; Tabata, S.
   #journal      DNA Res. (1996) 3:109-136
   #title        Sequence analysis of the genome of the unicellular
                   cyanobacterium Synechocystis sp. PCC6803. II. Sequence
                   determination of the entire genome and assignment of
                   potential protein-coding regions.
   #cross-references MUID:97061201
   #accession    S77548
      ##status        nucleic acid sequence not shown; translation not shown
      ##molecule_type DNA
      ##residues      1-241 ##label KAN
      ##cross-references EMBL:D90905; GB:AB001339; NID:g1652360; PID:d1018128;
                        PID:g1652473
      ##note          the nucleotide sequence was submitted to the EMBL Data
                       Library, June 1996
CLASSIFICATION   #superfamily response regulator homology
KEYWORDS         phosphoprotein
FEATURE
   5-115             #domain response regulator homology #label RRH\
   54                #binding_site phosphate (Asp) (covalent) #status
```

```
                    predicted
SUMMARY          #length 241  #molecular-weight 27405  #checksum 8759

  Query Match            14.6%;  Score 83;  DB 2;  Length 241;
  Best Local Similarity  28.2%;  Pred. No. 2.67e+00;
  Matches   43;  Conservative   30;  Mismatches  40;  Indels   1;  Gaps   1;

Db      81 LIAFSDGFTLERAQKINPQGYVIKPFRREQLLSI 114
           | |  ::  |:|||: :| ::  |:|   |||:
Qy      33 LLAATFTSTIERAHNFSP-SYIRHPYRPGQLLSS 65


RESULT   14
ENTRY            S49623      #type complete
TITLE            crtC protein - Rhodobacter sphaeroides
ORGANISM         #formal_name Rhodobacter sphaeroides
DATE             05-Mar-1995 #sequence_revision 12-May-1995 #text_change
                   02-Jul-1998
ACCESSIONS       S49623; S23634
REFERENCE        S49619
   #authors      Lang, H.P.; Coodell, R.J.; Takaichi, S.; Hunter, C.N.
   #submission   submitted to the EMBL Data Library, November 1994
   #description  The complete DNA sequence, specific TNS insertion map and
                   gene assignment of the carotenoid biosynthesis genes of
                   Rhodobacter sphaeroides.
   #accession    S49623
      ##status         preliminary
      ##molecule_type DNA
      ##residues       1-278 ##label LAN
      ##cross-references EMBL:X82458
REFERENCE        S23634
   #authors      Gari, E.; Toledo, J.C.; Gibert, I.; Barbe, J.
   #journal      FEMS Microbiol. Lett. (1992) 93:103-108
   #title        Nucleotide sequence of the methoxyneurosporene dehydrogenase
                   gene from Rhodobacter sphaeroides: comparison with other
                   bacterial carotenoid dehydrogenases.
   #accession    S23634
      ##status         preliminary; translation not shown
      ##molecule_type DNA
      ##residues       'M',2-42,'R',44 ##label GAR
      ##cross-references EMBL:X63204; NID:g46419; PID:g581532
GENETICS
   #gene         crtC
   #start_codon  GTG
SUMMARY          #length 278  #molecular-weight 31080  #checksum 4734

  Query Match            14.6%;  Score 83;  DB 2;  Length 278;
  Best Local Similarity  27.4%;  Pred. No. 2.67e+00;
  Matches   47;  Conservative   21;  Mismatches  19;  Indels   5;  Gaps   5;

Db      56 VGPS-RMHWTGTQLIVFVNFISS PPLVSPVKGRIVLTPTGITDVEVTLKDDGSHIWRPF 113
           :||   : || |||: :  : :| |||:   |  ::|:::|  : :| : :   | |::|:
Qy       1 MGPVFSLLWT-TQFFANARQSISLPPFI-IFKN-LLLAATFTSTIERAHNFSPSYIRHPY 57

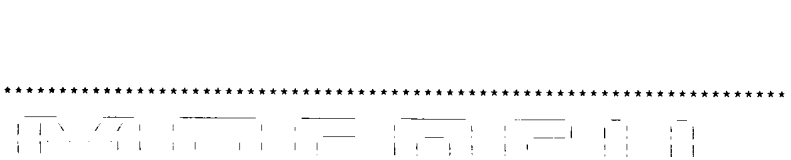
Db     114 AP 115
           |
Qy      58 RP 59


RESULT   15
ENTRY            A71216      #type complete
TITLE            hypothetical protein PH1993 - Pyrococcus horikoshii
ORGANISM         #formal_name Pyrococcus horikoshii
DATE             14-Aug 1998 #sequence_revision 14-Aug-1998 #text_change
                   14-Aug-1998
ACCESSIONS       A71216
REFERENCE        A71000
   #authors      Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.;
                   Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.;
                   Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura  K.; Otsuka, R.;
                   Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.;
                   Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi,
                   A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.;
                   Horikoshi, K.; Masuchi, Y.; Shizuya,
   #journal      DNA Res. (1998) 5:55-76
   #title        Complete sequence and gene organizatio
                   hyper-thermophilic archaebacterium,
                   0T3.
   #cross-references MUID:98344157
   #accession    A71216
      ##status         pre liminary; nucleic acid sequence
                         translation not shown
      ##molecule_type DNA
      ##residues       1-238 ##label KAW
      ##cross-references GB:AP000007; NID:g4236144; PID
      ##experimental_source strain ST
      ##note           this accession replaces an older
                         sequence replaced by GenBank
GENETICS
   #gene         PH1993
SUMMARY          #length 238  #molecular-weight 27501

  Query Match            14.3%;  Score 82;  DB 2;  Leng
  Best Local Similarity  41.0%;  Pred. No. 4.65e+00;
  Matches   23;  Conservative    5;  Mismatches  11;

Db      14 LPPFVPFPMILLLLNGNPLS:::HHVSPVY 44
           |||||:|: ||| :   :  :   :  : |
Qy      23 LPPFLLFKNLLLAATFTST:::AHNFSPSY 52


Search completed: Fri Aug 25 17:48:38 2000
Job time : 24 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 25 17:48:55 2000; MasPar time 4.52 Seconds
528.209 Million cell updates/sec

Tabular output not generated.

Title: >US-09-257-585-3
Description: (1-77) from US09257585.pep
Perfect Score: 568
Sequence: 1 MGPVPSLLWTTQFFANARQS ...RPCGLLSSSQCLERLHHPTL 77

Scoring table: PAM 150
Gap 11

Searched: 85661 seqs, 30989116 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 38.911; Variance 63.226; scale 0.615

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	103	18.1	971	1	MCM3_YEAST	MINICHROMOSOME MAINTEN	4.02e-04
2	69	12.3	476	1	?ANA_PSEFL	ARABINOSE-BINDING PRO	2.86e-03
3	91	16.0	382	1	NDF2_HUMAN	NEUROGENIC DIFFERENTIA	4.05e-02
4	91	16.0	383	1	NDF2_MOUSE	NEUROGENIC DIFFERENTIA	4.05e-02
5	89	15.7	382	1	NDF2_RAT	NEUROGENIC DIFFERENTIA	8.46e-02
6	87	15.3	221	1	RNC_AQUAE	RIBONUCLEASE III (EC 3	1.74e-01
7	84	14.8	389	1	EI_ORYSA	PUTATIVE TRANSCRIPTION	5.06e-01
8	83	14.6	162	1	MRP4_HUMAN	MULTIDRUG RESISTANCE A	7.18e-01
9	83	14.6	230	1	C552_THIFE	CYTOCHROME C-552 PRECU	7.18e-01
10	83	14.6	278	1	CRIC_RHOSH	HYDROXYNEUROSPORENE SY	7.18e-01
11	82	14.4	240	1	IBP6_HUMAN	INSULIN LIKE GROWTH FA	1.02e+00
12	82	14.4	1298	1	ICP4_HSV11	TRANS ACTING TRANSCRIP	1.02e+00
13	81	14.3	207	1	RASL_CAEEL	RAS-LIKE PROTEIN F54C8	1.43e+00
14	81	14.3	758	1	YLDM_CAEEL	HYPOTHETICAL 85.0 KDA	1.43e+00
15	80	14.1	541	1	CGT_RAT	2-HYDROXYACYLSPHINGOSI	2.01e+00
16	80	14.1	541	1	CGT_MOUSE	2-HYDROXYACYLSPHINGOSI	2.01e+00
17	80	14.1	637	1	NU5M_STRPU	NADH-UBIQUINONE OXIDOR	2.01e+00
18	80	14.1	644	1	UROM_RAT	UROMODULIN PRECURSOR (	2.01e+00
19	80	14.1	690	1	HRPO_BURSO	HYPERSENSITIVITY RESPO	2.01e+00
20	80	14.1	693	1	RECQ_ECOLI	ATP-DEPENDENT DNA HELI	2.01e+00
21	79	13.9	156	1	RS7_SPIPL	30S RIBOSOMAL PROTEIN	2.81e+00
22	79	13.9	341	1	IM37_SPIOL	37 KDA INNER ENVELOPE	2.81e+00
23	79	13.9	555	1	RT04_BRANA	MITOCHONDRIAL RIBOSOMA	2.81e+00
24	79	13.9	1164	1	Y_IA_HUMAN	MHC CLASS I	2.81e+00
25	78	13.7	375	1	AROB_MYCTU	3-DEHYDROQU	3.94e+00
26	78	13.7	541	1	HEX2_MOUSE	B(+)/HEX-SE	3.94e+00
27	78	13.7	541	1	HEX2_HUMAN	2 INTRACATA	3.94e+00
28	78	13.7	658	1	GAG_SEV1	GAG POLYPRO	3.94e+00
29	78	13.7	808	1	VGB3_NPVBC	VACCINIA AS	3.94e+00
30	78	13.7	1344	1	EIF3A_MOUSE	EUKARYOTIC	3.94e+00
31	78	13.7	1382	1	EIF3A_HUMAN	EUKARYOTIC	3.94e+00
32	77	13.6	517	1	QOX1_SULAC	QUINOL OXID	5.14e+00
33	77	13.6	1527	1	MRP3_HUMAN	CANALICULAR	5.14e+00
34	77	13.6	2947	1	FAPX_HUMAN	PROBABLE UB	5.14e+00
35	77	13.6	2959	1	FAPX_MOUSE	PROBABLE UB	5.14e+00
36	76	13.4	141	1	UL41_HCMVA	HYPOTHETICA	6.71e+00
37	76	13.4	156	1	RS7_ANANI	30S RIBOSOM	6.71e+00
38	76	13.4	178	1	PGSA_AQUAE	CDP-DIACYLG	6.71e+00
39	76	13.4	423	1	GRB4_CAEEL	GLYCINE REC	6.71e+00
40	76	13.4	620	1	VLCS_HUMAN	VERY-LONG-C	6.71e+00
41	76	13.4	645	1	HRPC2_XANCV	HRPC2 PROTE	6.71e+00
42	76	13.4	657	1	GAG_MPMV	GAG POLYPRO	6.71e+00
43	76	13.4	909	1	CUTA_FUSSO	CUTINASE PR	6.71e+00
44	76	13.4	923	1	YAUA_SCHPO	HYPOTHETICA	6.71e+00
45	76	13.4	1561	1	MRP1_HUMAN	MULTIDRUG R	6.71e+00

ALIGNMENTS

RESULT 1
ID MCM3_YEAST STANDARD; PRT; 971 AA.
AC P24279;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE MINICHROMOSOME MAINTENANCE PROTEIN 3.
GN MCM3 OR YEL032W OR SYGP-ORF23.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Sac
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91042499.
RA Gibson S.I., Surosky R.T., Tye B.K.;
RT "The phenotype of the minichromosome maintenance m
RT characteristic of mutants defective in DNA replica
RL Mol. Cell. Biol. 10:5707-5720 (1990).
RN [2]
RP SEQUENCE FROM N.A.
RA Mulligan J.T., Dietrich F.S., Bennessey K.M., Sell
RA Wei Y., Taylor P., Nakahara K., Roberts D., Davis
RL Submitted (FEB-1994) to the EMBL/GenBank/DDBJ data
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Dietrich F.S., Mulligan J.T., Bennessey K.M., Alle
RA Aviles E., Berno A., Brennan T., Carpenter J., Che
RA Chung E., Duncan M., Guzman E., Hartzell G., Hanle
RA Hyman R., Kaysser A., Komp C., Lashkari D., Lew H.
RA Mosedale D., Nakahara K., Namath A., Norgren R., P
RA Petel F.X., Roberts D., Sehl P., Schramm S., Shogr
RA Taylor P., Wei Y., Yelton M., Botstein D., Davis R
RL Submitted (FEB-1994) to the EMBL/GenBank/DDBJ data
CC ! FUNCTION: NECESSARY FOR CELL GROWTH. POSSIBLY
CC ROLE IN THE INITIATION OF CHROMOSOMAL DNA REPL
CC INTERACTING DIRECTLY WITH AUTONOMOUSLY REPLICA
CC (ARS). ITS ARS SPECIFICITY IS NOT IDENTICAL TO
CC MCM2 PROTEIN.
CC ! SUBCELLULAR LOCATION: N. (NUCLEAR).
CC -!- SIMILARITY: BELONGS TO THE MCM FAMILY.
CC ---
CC This SWISS-PROT entry is copyright. It is produced
CC between the Swiss Institute of Bioinformatics and
CC the European Bioinformatics Institute. There are
CC use by non-profit institutions as long as its
CC modified and this statement is not removed. Usage

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X53540; CAA37616.1; -.
DR   EMBL; U18779; AAB65010.1; -.
DR   PIR; A36376; A36376.
DR   PIR; S30851; S30851.
DR   SGD; L0001039; MCM3.
DR   PFAM; PF00493; MCM; 1.
DR   PROSITE; PS00847; MCM_1; 1.
DR   PROSITE; PS50051; MCM_2; 1.
KW   Transcription regulation; DNA binding, Nuclear protein;
KW   DNA replication; ATP-binding.
FT   DOMAIN      359    566       MCM.
FT   NP_BIND     409    416       ATP (POTENTIAL).
SQ   SEQUENCE   971 AA;  107517 MW;  43DD4DACAF4456DC CRC64;

  Query Match           18.1%;  Score 103;  DB 1;  Length 971;
  Best Local Similarity  31.7%;  Pred. No. 4.02e-04;
  Matches   19;  Conservative   11;  Mismatches   25;  Indels   2;  Gaps   2;

Db     528 DPHQNIALPDSLLSPFDLLPVVTDDINEIPDRSISEHVLRTHRYLPPGYLFGEPVPERLN 587
           :::|: :|| || : :||: |:    | :: |  :| | | | | | :   |||:
Qy      16 NARQSTSLPPFLLFK-NLLLAAIEISTIERAHNESPSYIR-RPVRPCGLLSSSQCLFPLH 73


RESULT    2
ID   GANA_PSEFL     STANDARD;      PRT;   376 AA.
AC   P48841;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   ARABINOGALACTAN ENDO-1,4-BETA-GALACTOSIDASE PRECURSOR (EC 3.2.1.89)
DE   (ENDO-1,4-BETA-GALACTANASE) (GALACTANASE).
GN   GALA OR GANA.
OS   Pseudomonas fluorescens.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC   Pseudomonas.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SP. CELLULOSA;
RX   MEDLINE; 98060803.
RA   Braithwaite K.L., Barna T., Spurway T.D., Charnock S.J., Black G.W.,
RA   Hughes N., Lakey J.H., Virden R., Hazlewood G.P., Henrissat B.,
RA   Gilbert H.J.;
RT   "Evidence that galactanase A from Pseudomonas fluorescens subspecies
RT   cellulosa is a retaining family 53 glycosyl hydrolase in which E161
RT   and E270 are the catalytic residues.",
RL   Biochemistry 36:15489-15500(1997).
CC   -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GALACTOSIDIC
CC       LINKAGES IN ARABINOGALACTANS.
CC   -!- SIMILARITY: BELONGS TO FAMILY 53 OF GLYCOSYL HYDROLASES.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X91885; CAA62990.1; -.
KW   Hydrolase; Glycosidase; Signal.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18    376       ARABINOGALACTAN ENDO-1,4-BETA-
FT                                GALACTOSIDASE.
FT   ACT_SITE     161    161      PROTON DONOR.
FT   ACT_SITE     270    270      NUCLEOPHILE.
FT   MUTAGEN      161    161      E->A: PARTIAL LOSS OF ACTIVITY.
FT   MUTAGEN      270    270      E->A: COMPLETE LOSS OF ACTIVITY.
SQ   SEQUENCE   376 AA;  42315 MW;  BEC4F7243C6CAE10 CRC64;

  Query Match           17.3%;  Score 98;  DB 1;  Length 376;
  Best Local Similarity  43.5%;  Pred. No. 2.85e-04;
  Matches   10;  Conservative    6;  Mismatches   6;  Indels   1;  Gaps   1;

Db     344 WENASFFDATRKNNALPAFLFFK 366
           |  : ||  :| : :|||:||:||
Qy       9 W-TTQFFANARQSTSLPPFLLFK 30


RESULT    3
ID   NDF2_HUMAN     STANDARD;      PRT;   382 AA.
AC   Q15784;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15 FEB 2000 (Rel. 39, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   NEUROGENIC DIFFERENTIATION FACTOR 2.
GN   NEUROD2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata, Vertebrata, Euteleostomi,
OC   Mammalia, Eutheria, Primates, Catarrhini, Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE, 96413331.
RA   McCormick M.B., Tamimi R.M., Snider L., Asakura A., Bergstrom D.,
RA   Tapscott S.J.;
RT   "NeuroD2 and neuroD3: distinct expression patterns and
RT   transcriptional activation potentials within the neuroD gene
RT   family.";
RL   Mol. Cell. Biol. 16:5792-5800(1996).
RN   [2]
RP   REVISIONS TO 336-338.
RA   Tapscott S.J., Tamimi R.T., McCormick B.M.;
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION. APPEARS TO MEDIATE NEURONAL DIFFERENTIATION.
CC   -!- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC       BHLH PROTEIN.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC   -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC       TRANSCRIPTION FACTORS. "ATONAL" SUBFAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U58681; AAC26058.1; -.
DR   HSSP; P10085; 1MDY.
DR   MIM; 601725; -.
DR   PFAM; PF00010; HLH; 1.
DR   PROSITE; PS00038; HELIX_LOOP_HELIX; 1.
KW   Transcription regulation; DNA-binding; Nuclear protein;
KW   Differentiation.
FT   DOMAIN       37     46       POLY-PRO.
FT   DOMAIN       82     91       POLY-GLU.
FT   DOMAIN      107    113       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DNA_BIND    122    133       BASIC DOMAIN.
FT   DOMAIN      134    174       HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT   DOMAIN      282    285       POLY-GLY.
SQ   SEQUENCE   382 AA;  41361 MW;  E33AA5C0BE371B47 CRC64;

  Query Match           16.0%;  Score 91;  DB 1;  Length 382;
  Best Local Similarity  29.7%;  Pred. No. 4.05e-02;
  Matches   11;  Conservative   10;  Mismatches   15;  Indels   1;  Gaps   1;

Db     215 FLTEQCADGAGRFHGSGGPEAMHPYPYPGSRLACAQC 251
           .|.   .    | |  : ::  |||  ||: |:::||
Qy      33 LLAATETSTIERAHNESPSYIRHPYP-PCGLLSSSQC 68


RESULT    4
ID   NDF2_MOUSE     STANDARD;      PRT;   383 AA.
```

```
AC   Q62414; Q61952;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-JUL-1999 (Rel. 48, Last sequence update)
DT   15-JUL-1999 (Rel. 48, Last annotation update)
DE   NEUROGENIC DIFFERENTIATION FACTOR 2 (NEUROD RELATED FACTOR) (NDRF).
GN   NEUROD2 OR NDRF.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96413331.
RA   McCormick M.B., Tamimi R.M., Snider L., Asakura A., Bergstrom D.,
RA   Tapscott S.J.;
RT   "NeuroD2 and neuroD3: distinct expression patterns and
RT   transcriptional activation potentials within the neuroD gene
RT   family.";
RL   Mol. Cell. Biol. 16:5792-5800(1996).
RN   [2]
RP   REVISIONS TO 348-349.
RA   Tapscott S.J., McCormick M.B., Tamimi R.T.;
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE; 96183021.
RA   Yasunami M., Suzuki K., Maruyama H., Kawakami H., Nagai Y.,
RA   Hagiwara M., Ohkubo H.;
RT   "Molecular cloning and characterization of a cDNA encoding a novel
RT   basic helix-loop-helix protein structurally related to
RT   Neuro-D/BHF1.";
RL   Biochem. Biophys. Res. Commun. 220:754-758(1996).
CC   -!- FUNCTION: APPEARS TO MEDIATE NEURONAL DIFFERENTIATION.
CC   -!- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC       BHLH PROTEIN.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC   -!- TISSUE SPECIFICITY: APPEARS IN THE BRAIN OF 12-DAY-OLD MOUSE
CC       EMBRYOS AND IS LOCALIZED IN CERTAIN REGIONS OF THE ADULT BRAIN,
CC       SUCH AS THE HIPPOCAMPUS, DENTATE GYRUS AND CEREBELLUM.
CC   -!- DEVELOPMENTAL STAGE: INITIALLY EXPRESSED AT EMBRYONIC DAY 11, WITH
CC       PERSISTENT EXPRESSION IN THE ADULT NERVOUS SYSTEM.
CC   -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC       TRANSCRIPTION FACTORS. "ATONAL" SUBFAMILY.
CC   ---------- ------------------------------ ------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ------------------------------ ------------------------------
DR   EMBL; U58471; AAC26057.1; -.
DR   EMBL; D83507; BAA11931.1; -.
DR   HSSP; P10085; 1MDY.
DR   MGD; MGI:107755; NEUROD2.
DR   PFAM; PF00010; HLH; 1.
DR   PROSITE; PS00038; HELIX_LOOP_HELIX; 1.
KW   Transcription regulation; DNA-binding; Nuclear protein;
KW   Differentiation.
FT   DOMAIN       83     92       POLY-GLU.
FT   DNA_BIND     123    134      BASIC DOMAIN.
FT   DOMAIN       135    175      HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT   DOMAIN       283    286      POLY-GLY.
FT   DOMAIN       2      2        NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   CONFLICT     41     41       S -> P (IN REF. 3).
SQ   SEQUENCE   383 AA;  41483 MW;  EB18DDF6336A35F0 CRC64;

  Query Match           16.0%;  Score 91;  DB 1;  Length 383;
  Best Local Similarity  29.7%;  Pred. No. 4.05e-02;
  Matches   11;  Conservative   10;  Mismatches  15;  Indels   1;  Gaps   1;

Db    216 FLTEQGADGAGKFHQSGCPFAMHPYPYPCSKLAGAQC 252
```

```
Qy     33 LLAATEISTIERAHNEPSYFSDAP PGGLLSSSQC 68


RESULT    5
ID   NDF2_RAT        STANDARD;      PRT;   362 AA.
AC   Q63689;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   NEUROGENIC DIFFERENTIATION FACTOR 2 (BRAIN BHLH PR
GN   NEUROD2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrat
OC   Mammalia; Eutheria; Rodentia; Sciuromorpha; Murid
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN WISTAR; TISSUE=BRAIN;
RX   MEDLINE; 96193685.
RA   Kume H., Maruyama K., Endo T., Iwatsubo T., Sai
RT   "Molecular cloning of a novel basic helix-loop-he
RT   rat brain.";
RL   Biochem. Biophys. Res. Commun. 219:526-530(1996).
RN   [2]
RP   CONCEPTUAL TRANSLATION.
RA   Hulo C.;
RL   Unpublished observations (AUG-1998).
CC   -!- FUNCTION: APPEARS TO MEDIATE NEURONAL DIFFERE
CC   -!- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMER
CC       BHLH PROTEIN.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC   -!- TISSUE SPECIFICITY: DETECTED ONLY IN NEURAL
CC   -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-H
CC       TRANSCRIPTION FACTORS, "ATONAL" SUBFAMILY.
CC   -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION W
CC       INTRODUCED IN POSITIONS 225 AND 250 TO EXTEND
CC       THE HUMAN AND MOUSE ORTHOLOGS.
CC   ---------- ------------------------------ ------
CC   This SWISS-PROT entry is copyright. It is produce
CC   between the Swiss Institute of Bioinformatics a
CC   the European Bioinformatics Institute. There are
CC   use by non-profit institutions as long as its
CC   modified and this statement is not removed. Usag
CC   entities requires a license agreement (See http:/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ------------------------------ ------
DR   EMBL; D82868; BAA11610.1; ALT_FRAME.
DR   PFAM; PF00010; HLH; 1.
DR   PROSITE; PS00038; HELIX_LOOP_HELIX; 1.
KW   Transcription regulation; DNA-binding; Nuclear p
KW   Differentiation.
FT   DOMAIN       82     91       POLY-GLU.
FT   DNA_BIND     122    133      BASIC DOMAIN.
FT   DOMAIN       134    174      HELIX-LOOP-HELIX MO
FT   DOMAIN       282    285      POLY-GLY.
SQ   SEQUENCE   362 AA;  41468 MW;  05234DE9BE0A8C09

  Query Match           15.2%;  Score 89;  DB 1;  Le
  Best Local Similarity  29.7%;  Pred. No. 8.46e-02;
  Matches   11;  Conservative   10;  Mismatches  15;

Db    215 FLTEQGADGAGKFHQSGCPFAMHPYPYPCSKLAGDQC 25
Qy     33 LLAATEISTIERAHNEPSYFSDAP PGGLLSSSQC 68


RESULT    6
ID   RNC_AQUAE       STANDARD;      PRT;   221 AA.
AC   O67082;
DT   15-FEB-2000 (Rel. 39, Created)
DT   15-FEB-2000 (Rel. 39, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   RIBONUCLEASE III (EC 3.1.26.3) (RNASE III).
```

```
GN   RNC OR AQ_946.
OS   Aquifex aeolicus.
OC   Bacteria; Aquificales; Aquificaceae; Aquifex.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VF5;
RX   MEDLINE; 98196666.
RA   Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA   Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA   Feldman R.A., Short J.M., Olsen G.J., Swanson R.V.;
RT   "The complete genome of the hyperthermophilic bacterium Aquifex
RT   aeolicus.";
RL   Nature 392:353-358(1998).
CC   -!- FUNCTION: DIGESTS DOUBLE-STRANDED RNA. INVOLVED IN THE PROCESSING
CC       OF RIBOSOMAL RNA PRECURSORS AND OF SOME MRNAS (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ENDONUCLEOLYTIC CLEAVAGE TO 5'-PHOSPHO-
CC       MONOESTER.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- SIMILARITY: CONTAINS 1 DRBM (DOUBLE-STRANDED RNA-BINDING) DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE RNASE III FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000715; AAC07049.1; -.
DR   PFAM; PF00035; dsrm; 1.
DR   PFAM; PF00636; Ribonuclease_3; 1.
DR   PROSITE; PS00517; RNASE_3_1; 1.
KW   Hydrolase; Nuclease; Endonuclease; RNA-binding.
FT   DOMAIN      201    217       DRBM.
SQ   SEQUENCE   221 AA;  26100 MW;  ACB6073508A4C611 CRC64;

  Query Match           15.3%;  Score 87;  DB 1;  Length 221;
  Best Local Similarity  24.5%;  Pred. No. 1.74e-01;
  Matches   12;  Conservative   17;  Mismatches  18;  Indels   2;  Gaps   2;

Db     105 IGDVFEALWAAVYIDSCRDANFTRELFYKLFKEDILSAIKRGRVKKDYK 153
           :| ||  |||: :: ::|::  :   |  |||: :|:|   : : :  :
Qy       1 MGPVFSLLWTTQFFANARQST-SLPPFL-LFKNLLLAATFTSTIERAHN 47


RESULT    7
ID   FL_ORYSA       STANDARD;      PRT;   389 AA.
AC   O21175;
DT   15-FEB-2000 (Rel. 39, Created)
DT   15-FEB-2000 (Rel. 39, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   PUTATIVE TRANSCRIPTION FACTOR FL (RFL).
GN   FL.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. TORIDE; TISSUE=PANICLE;
RA   Kyozuka J., Konishi S., Nemoto K., Izawa T., Shimamoto K.;
RT   "Downregulation of RFL, rice FLO/LFY homolog, accompanied with panicle
RT   branching.";
RL   Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PROBABLE TRANSCRIPTION FACTOR (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC   -!- SIMILARITY: BELONGS TO THE FLO / LFY FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB005620; BAA21547.1; -.
DR   MENDEL; 25648; Orysa;1429;25648.
DR   PFAM; PF01698; FLO_LFY; 1.
KW   Transcription regulation; Activator; DNA-binding;
KW   Nuclear protein; Developmental protein.
FT   DOMAIN       22     37       POLY-PRO.
FT   DOMAIN      201    204       POLY-GLY.
FT   DOMAIN      213    217       POLY-GLY.
FT   DOMAIN      378    383       POLY-ALA.
SQ   SEQUENCE   389 AA;  42530 MW;  5172AA6A34B0868A CRC64;

  Query Match           14.8%;  Score 84;  DB 1;  Length 389;
  Best Local Similarity  31.7%;  Pred. No. 5.06e-01;
  Matches   13;  Conservative    9;  Mismatches  18;  Indels   1;  Gaps   1;

Db     225 PVVTEPGEVAPAKKNCLDYLPHLYEQCELFLLQVGSMAKLR 265
           :  || : : || :::  |: | |  | | |   | : :||
Qy      34 LAATFTSTIEPAHNESPSYIRHPYRPCGL-LSSSQCLERLR 73


RESULT    8
ID   MRP4_HUMAN     STANDARD;      PRT;   162 AA.
AC   O15439;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   MULTIDRUG RESISTANCE-ASSOCIATED PROTEIN 4 (FRAGMENT).
GN   ABCC4 OR MRP4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 97413640.
RA   Kool M., de Haas M., Scheffer G.L., Scheper R.J., van Eijk M.J.,
RA   Juijn J.A., Baas F., Borst P.;
RT   "Analysis of expression of cMOAT (MRP2), MRP3, MRP4, and MRP5,
RT   homologues of the multidrug resistance associated protein gene
RT   (MRP1), in human cancer cell lines.";
RL   Cancer Res. 57:3537-3547(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RA   Kool M., de Haas M., Ponne N.J., Baas F., Borst P.;
RL   Submitted (JAN 1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC       (ABC TRANSPORTERS). MRP SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U83660; AAB71757.1; -.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW   ATP binding; Glycoprotein; Transmembrane; Transport; Repeat.
FT   NON_TER       1      1
FT   CARBOHYD     22     22       POTENTIAL.
FT   NON_TER     162    162
SQ   SEQUENCE   162 AA;  18411 MW;  CB080153C33D246A CRC64;

  Query Match           14.6%;  Score 84;  DB 1;  Length 162;
  Best Local Similarity  25.8%;  Pred. No. 7.18e-01;
  Matches   16;  Conservative   15;  Mismatches  29;  Indels   2;  Gaps   2;
```

```
Db      21 SNFSVSQRQLVCLARATLRKNQTLTTDEAIANVDPRTDRLTQKKTREKFAHCIVLTIAHR 80
           :|    ||   |: :| || ::   | |: :: |: :     ||  :  | :|: ::
Qy      11 TQFFANARQSTSLPPFLIFFNILLAAIE IFTIE RAHNESPSYIRHFYRLSLFSSQC 68

Db      81 LN 82
           |:
Qy      69 LE 70


RESULT    9
ID   C552_THIFE     STANDARD;      PRT:   230 AA.
AC   P74977;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   CYTOCHROME C-552 PRECURSOR (C552).
GN   CYC1.
OS   Thiobacillus ferrooxidans.
OC   Bacteria; Proteobacteria; gamma subdivision; Acidithiobacillus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 33020;
RA   Appia C., Bengrine A., Chippaux M., Cavazza C., Bruschi M.,
RA   Bonnefoy V.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 48-95, AND CHARACTERIZATION.
RX   MEDLINE; 97129108.
RA   Cavazza C., Giudici-Orticoni M.-T., Nitschke W., Appia C.,
RA   Bonnefoy V., Bruschi M.;
RT   "Characterisation of a soluble cytochrome c4 isolated from
RT   Thiobacillus ferrooxidans.";
RL   Eur. J. Biochem. 242:308-314(1996).
CC   -!- FUNCTION: DIHEME, HIGH POTENTIAL CYTOCHROME C WITH REDOX
CC       POTENTIALS OF 485 MV (+-20 MV) AND 480 MV (+-20 MV).
CC   -!- SUBCELLULAR LOCATION: PERIPLASMIC.
CC   -!- PIM: BINDS TWO HEME GROUPS PER MOLECULE.
CC   -!- MASS SPECTROMETRY: MW=21193; METHOD=ELECTROSPRAY; RANGE=30-213.
CC   -!- SIMILARITY: BELONGS TO THE CLASS C4 FAMILY OF DIHEME CYTOCHROMES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y07610; CAA68880.1; -.
DR   HSSP; Q52369; 1ETP.
DR   PFAM; PF00034; cytochrome_c; 2.
DR   PROSITE; PS00190; CYTOCHROME_C; 2.
KW   Electron transport; Heme; Periplasmic; Signal.
FT   SIGNAL        1     47
FT   CHAIN        48    230       CYTOCHROME C-552.
FT   BINDING      63     63       HEME 1 (COVALENT) (BY SIMILARITY).
FT   BINDING      66     66       HEME 1 (COVALENT) (BY SIMILARITY).
FT   METAL        67     67       IRON 1 (HEME AXIAL LIGAND)
FT                                (BY SIMILARITY).
FT   BINDING     166    166       HEME 2 (COVALENT) (BY SIMILARITY).
FT   BINDING     169    169       HEME 2 (COVALENT) (BY SIMILARITY).
FT   METAL       170    170       IRON 2 (HEME AXIAL LIGAND)
FT                                (BY SIMILARITY).
FT   CONFLICT     89     89       Y -> YI (IN REF. 2).
SQ   SEQUENCE   230 AA;  25398 MW;  B762287CF7B1AF57 CRC64;

  Query Match           14.6%;  Score 83;  DB 1;  Length 230;
  Best Local Similarity  37.0%;  Pred. No. 7.18e-01;
  Matches   10;  Conservative    8;  Mismatches   9;  Indels   0;  Gaps   0;

Db      31 FLLFSALLIVAGQASAAVGSADAPAPY 57
           |||| ||| | ::|:  : : :::|
Qy      26 FLLFKNLLLAAIEISLIERAHNESPSY 52


RESULT   10
ID   CRTC_RHOSH     STANDARD;      PRT:   278 AA.
AC   Q01670;
DT   01-JUL-1993 (Rel. 29, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   HYDROXYNEUROSPORENE DEHYDROGENASE (EC 1. . . .) (HY
DE   SYNTHASE).
GN   CRTC.
OS   Rhodobacter sphaeroides (ex Pseudomonas sphaer
OC   Bacteria; Proteobacteria; alpha subdivision; Rhod
OC   Rhodobacter.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;
RX   MEDLINE; 95238278.
RA   Lang H.P., Cogdell R.J., Takaichi S., Hunter C.N.
RT   "Complete DNA sequence, specific Tn5 insertion map
RT   assignment of the carotenoid biosynthesis pathway
RT   sphaeroides.";
RL   J. Bacteriol. 177:2064-2073(1995).
RN   [2]
RP   SEQUENCE OF 1-44 FROM N.A.
RC   STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;
RX   MEDLINE; 92307398.
RA   Gari E., Toledo J.C., Gibert I., Barbe J.;
RT   "Nucleotide sequence of the methoxyneurosporene-d
RT   from Rhodobacter sphaeroides: comparison with othe
RT   carotenoid dehydrogenases.";
RL   FEMS Microbiol. Lett. 72:255-260(1992).
CC   -!- PATHWAY: CAROTENOID AND CHLOROPHYLL BIOSYNTHES
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced
CC   between the Swiss Institute of Bioinformatics an
CC   the European Bioinformatics Institute. There are
CC   use by non-profit institutions as long as its
CC   modified and this statement is not removed. Usage
CC   entities requires a license agreement (See http:/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ010302; CAB
DR   EMBL; X63204; CAA44
KW   Photosynthesis; Chloroplast; Biosynthesis; Carote
FW   Oxidoreductase.
FT   CONFLICT     43     43       A -> P (IN REF. 2).
SQ   SEQUENCE   278 AA;  MW:  4F48E46F72B8B711 CR

  Query Match           14.6%;  Score 83;  DB 1;  Leng
  Best Local Similarity  27.4%;  Pred. No. 7.18e-01;
  Matches   17;  Conservative   21;  Mismatches   19;

Db      56 VGPS-RMHWIGTQEIVEVNEISS PPGVSPVKGRIVLIPGTIDA
           :||   : || ||:: :  : :: ::|  |  ::|:  :  :
Qy       1 MCPVFSLLWI TQFLANAGGS GPLL LFKN LLLAAIFISI

Db     114 AP 115
           |
Qy      58 RP 59


RESULT   11
ID   IBP6_HUMAN     STANDARD;      PRT:   240 AA.
AC   P24592; Q14492;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN 6 PRECU
DE   (IBP-6) (IGF-BINDING PROTEIN 6).
GN   IGFBP6 OR IBP6.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata
```

```
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=OSTEOSARCOMA;
RX   MEDLINE; 91225006.
RA   Kiefer M.C., Masiarz F.R., Bauer D.M., Zapf J.;
RT   "Identification and molecular cloning of two new 46 kDa insulin-like
RT   growth factor binding proteins isolated from adult human serum.";
RL   J. Biol. Chem. 266:9043-9049(1991).
RN   [2]
RP   SEQUENCE OF 2-240 FROM N.A.
RC   TISSUE=PLACENTA;
RX   MEDLINE; 92049376.
RA   Shimasaki S., Gao L., Shimonaka M., Ling N.;
RT   "Isolation and molecular cloning of insulin-like growth
RT   factor-binding protein-6.";
RL   Mol. Endocrinol. 5:938-948(1991).
RN   [3]
RP   PARTIAL SEQUENCE OF 28-42.
RX   MEDLINE; 91207395.
RA   Andress D.L., Birnbaum R.S.;
RT   "A novel human insulin-like growth factor binding protein secreted by
RT   osteoblast-like cells.";
RL   Biochem. Biophys. Res. Commun. 176:213-218(1991).
RN   [4]
RP   PRELIMINARY SEQUENCE OF 28-42.
RC   TISSUE=CEREBROSPINAL FLUID;
RX   MEDLINE; 90005986.
RA   Roghani M., Hossenlopp P., Lepage P., Bailand A., Binoux M.;
RT   "Isolation from human cerebrospinal fluid of a new insulin-like
RT   growth factor-binding protein with a selective affinity for IGF-II.";
RL   FEBS Lett. 255:253-258(1989).
RN   [5]
RP   CARBOHYDRATE-BINDING SITES.
RX   MEDLINE; 98241382.
RA   Neumann G.M., Marinaro J.A., Bach L.A.;
RT   "Identification of O-gl,Oss,lation sites and partial characterization
RT   of carbohydrate structure and disulfide linkages of human
RT   insulin-like growth factor binding protein 6.";
RL   Biochemistry 37:6572-6585(1998).
RN   [6]
RP   DISULFIDE BONDS.
RX   MEDLINE; 99262603.
RA   Neumann G.M., Bach L.A.;
RT   "The N-terminal disulfide linkages of human insulin-like growth
RT   factor-binding protein-6 (hIGFBP-6) and hIGFBP-1 are different as
RT   determined by mass spectrometry.";
RL   J. Biol. Chem. 274:14587-14594(1999).
CC   -!- FUNCTION: IGF-BINDING PROTEINS PROLONG THE HALF-LIFE OF THE IGFS
CC       AND HAVE BEEN SHOWN TO EITHER INHIBIT OR STIMULATE THE GROWTH
CC       PROMOTING EFFECTS OF THE IGFS ON CELL CULTURE. THEY ALTER THE
CC       INTERACTION OF IGFS WITH THEIR CELL SURFACE RECEPTORS.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- PTM: O-GLYCOSYLATED.
CC   -!- SIMILARITY: CONTAINS 1 THYROGLOBULIN TYPE-I DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M62402; AAB06187.1; -.
DR   EMBL; M69054; AAA88070.1; -.
DR   PIR; PH0144; PH0144.
DR   PIR; S05699; S05699.
DR   PIR; A39842; A39842.
DR   MIM; 146735; -.
DR   PFAM; PF00219; IGFBP; 1.
DR   PFAM; PF00086; thyroglobulin_1; 1.
DR   PROSITE; PS00222; IGF_BINDING; FALSE_NEG.
DR   PROSITE; PS00484; THYROGLOBULIN_1; 1.
KW   Growth factor binding; Signal; Glycoprotein.
FT   SIGNAL        1     24       PROBABLE.
FT   CHAIN        25    240       INSULIN LIKE GROWTH FACTOR BINDING
FT                                 PROTEIN 6.
FT   DOMAIN      186    234       THYROGLOBULIN TYPE I.
FT   DISULFID     29     32
FT   DISULFID     40     44
FT   DISULFID     57     63
FT   DISULFID     71     84
FT   DISULFID     78    104
FT   DISULFID    163    190
FT   DISULFID    201    212
FT   DISULFID    214    234
FT   CARBOHYD    126    126
FT   CARBOHYD    144    144
FT   CARBOHYD    145    145
FT   CARBOHYD    146    146
FT   CARBOHYD    152    152
FT   CONFLICT      2      2       T -> C (IN REF. 2).
SQ   SEQUENCE   240 AA;  25322 MW;  2853082310025069 CRC64;

  Query Match          14.4%;  Score 82;  DB 1;  Length 240;
  Best Local Similarity;  50.0%;  Pred. No. 1.6e+00;
  Matches   11;  Conservative    6;  Mismatches   5;  Indels   0;  Gaps   0;

Db       7 LPPLLLLLAALLLAASPGCALAK 28
           |||:||:  |||||:  ::: |
Qy      23 LPFPLLPKNLLLAATETSTIER 44


RESULT   12
ID   ICP4_HSV11     STANDARD;      PRT;  1298 AA.
AC   P08392;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-AUG-1988 (Rel. 08, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   TRANS-ACTING TRANSCRIPTIONAL PROTEIN ICP4 (TRANSCRIPTIONAL ACTIVATOR
DE   IE175) (ALPHA-4 PROTEIN).
GN   ICP4 OR IE175 OR RS1.
OS   Herpes simplex virus (type 1 / strain 17).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Simplexvirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88274327.
RA   McGeoch D.J., Dalrymple M.A., Davison A.J., Dolan A., Frame M.C.,
RA   McNab D., Perry L.J., Scott J.E., Taylor P.;
RT   "The complete DNA sequence of the long unique region in the genome of
RT   herpes simplex virus type 1.";
RL   J. Gen. Virol. 69:1531-1574(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 86148504.
RA   McGeoch D.J., Dolan A., Donald S., Brauer D.H.K.;
RT   "Complete DNA sequence of the short repeat region in the genome of
RT   herpes simplex virus type 1.";
RL   Nucleic Acids Res. 14:1727-1745(1986).
RN   [3]
RP   DNA-BINDING DOMAIN.
RX   MEDLINE; 90174974.
RA   Wu C.-L., Wilcox K.W.;
RT   "Codons 262 to 490 from the herpes simplex virus ICP4 gene are
RT   sufficient to encode a sequence-specific DNA binding protein.";
RL   Nucleic Acids Res. 18:531-538(1990).
RN   [4]
RP   INFLUENCE OF PHOSPHORYLATION ON FUNCTION.
RX   MEDLINE; 91122047.
RA   Papavassiliou A.G., Wilcox K.W., Silverstein S.J.;
RT   "The interaction of ICP4 with cell/infected cell factors and its
RT   state of phosphorylation modulate differential recognition of leader
```

```
RT   sequences in herpes simplex virus DNA.";
RL   EMBO J. 10:397-406(1991).
CC   -!- FUNCTION: THIS IE PROTEIN IS A MULTIFUNCTIONAL PROTEIN CAPABLE
CC       OF MIGRATING TO THE NUCLEUS, BINDING TO DNA, TRANS-ACTIVATING
CC       OTHER VIRAL GENES, AND AUTOREGULATING ITS OWN SYNTHESIS. BINDING
CC       OF ICP4 TO ALPHA GENES CAUSES REPRESSION OF THEIR EXPRESSION.
CC       ICP4 ALSO ACTS AS A TRANSCRIPTIONAL ACTIVATOR OF BETA AND GAMMA
CC       GENES. ICP4 BINDS WITH HIGH AFFINITY TO THE SEQUENCE 5'-ATCGTC-3'.
CC   -!- SUBUNIT: HOMODIMER.
CC   -!- SUBCELLULAR LOCATION: NUCLEUS OF INFECTED CELLS.
CC   -!- PTM: IC4 IS ADP-RIBOSYLATED.
CC   -!- PTM: THE LONG STRETCH OF SER IS A MAJOR SITE OF PHOSPHORYLATION.
CC       ONLY THE PHOSPHORYLATED FORMS OF ICP4 ARE CAPABLE OF INTERACTING
CC       WITH BETA OR GAMMA GENES.
CC   -!- SIMILARITY: BELONGS TO THE ICP4/IE110/ICP0 FAMILY
CC   ---------- ----------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ----------------------------------------------------------
DR   EMBL; L00036; AAA96675.1; -.
DR   EMBL; L00036; AAA96688.1; -.
DR   EMBL; X14112; CAA32286.1; -.
DR   EMBL; X14112; CAA32278.1; -.
DR   EMBL; X06461; CAA29763.1; -.
DR   PIR; A23510; EDBE75.
DR   HSSP; P04002; 1WFA.
KW   Early protein; Transcription regulation; Trans-acting factor;
KW   DNA-binding; Phosphorylation; Nuclear protein; ADP ribosylation.
FT   DOMAIN      176    199       SER-RICH.
FT   DNA_BIND    262    490
SQ   SEQUENCE   1298 AA;  132843 MW;  4F32E04C95CA9344 CRC64;

  Query Match           14.4%;  Score 82;  DB 1;  Length 1298;
  Best Local Similarity  28.4%;  Pred. No. 1.02e+00;
  Matches   19;  Conservative   19;  Mismatches  25;  Indels   4;  Gaps   4;

Db   1000 LLSTRDLAFAGAVEFLGLLASAGDRRLIVVNT-VRACDWP-ADGPAVSPQLAYLACELLP 1057
            || | :: || | :  : |   :  | ::  |   : :::  :::|  | |:| :| ||:
Qy      7 LLWTTQF-FANARQSTSIPPFLIFKNLLIAATETSTIERAHNESPSYIR-FPYRPCGLLS 64

Db   1058 AVQCAVR 1064
            : || !
Qy     65 SSQCLER 71


RESULT   13
ID   RASL_CAEEL     STANDARD;      PRT:   207 AA.
AC   P34443;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 29, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   RAS LIKE PROTEIN F54C8.5.
GN   F54C8.5.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA   Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA   Johnston L., Jones M., Kershaw J., Kirsten J., Laisster N.,
RA   Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Sims M., Smaldon N., Smith A., Smith M., Sounhammer E., Staden R.,
RA   Sulston J., Thierry-Mieg J., Thomas K., Vaudin M.,
RA   Waterston R., Watson A., Weinstock L., Wilkinson J.,
RA   Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
CC   -!- SIMILARITY: BELONGS T  THE SMALL GTPASE SUPER
CC   ---------- ----------  -  ---  ------------
CC   This SWISS-PROT entry is copyright. It is produced
CC   between the Swiss Institute of Bioinformatics an
CC   the European Bioinformatics Institute. There are
CC   use by non-profit institutions as long as its
CC   modified and this statement is not removed. Usage
CC   entities requires a license agreement (See http://
CC   or send an email to license@isb-sib.ch).
CC   ---------- ----------  -  ---  ------------
DR   EMBL; Z22178; CAA80[illegible].1; -.
DR   PIR; S40747; S40747.
DR   HSSP; P10113; 1GUA.
DR   WORMPEP; F54C8.5; CE00191.
DR   PFAM; PF00071; ras; 1.
DR   PRINTS; PR00449; RASTRNSFRMNG.
KW   Hypothetical protein; GTP-binding; Prenylation; I
FT   NP_BIND      20     27       GTP (BY SIMILARITY).
FT   NP_BIND      67     71       GTP (BY SIMILARITY).
FT   NP_BIND     126    129       GTP (BY SIMILARITY).
FT   DOMAIN       42     58       EFFECTOR REGION (BY
FT   LIPID       204    204       FARNESYL (BY SIMILAR
SQ   SEQUENCE   207 AA;  23695 MW;  [illegible]

  Query Match           14.4%;  Score 81;  DB 1;  Len
  Best Local Similarity  28.4%;  Pred. No. 1.43e+00;
  Matches   15;  Conservative   16;  Mismatches  17;

Db   148 WDAKRVELIARDSNEVPLVLVGNKSDLEDERQVSVEEGRECAVVFDCKFIETSAGNESPIEEFENNSF
          | : |   |||| :  : |  |:::: :: : |  |:|:: :|
Qy     9 WTTQFFA-NARQSTSIPP-F LIFKNLLIAATETSTIERAHNESP

RESULT   14
ID   YLDM_CAEEL     STANDARD;      PRT:   758 AA.
AC   Q94225;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   HYPOTHETICAL 85.0 KDA PROTEIN F41D9.5 IN CHROMOSOM
GN   F41D9.5.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditi
OC   Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Fulton R.;
RL   Submitted (SEP 1996) to the EMBL/GenBank/DDBJ data
RN   [2]
RP   REVISIONS.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (SEP 1997) to the EMBL/GenBank/DDBJ data
CC   -!- FUNCTION: POSSIBLE SULFATE TRANSPORTER.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEI
CC   -!- SIMILARITY: BELONGS TO THE SULFATE PERMEASE FA
CC   ---------- ----------  -  ---  ------------
CC   This SWISS-PROT entry is copyright. It is prod
CC   between the Swiss Institute of Bioinformatics an
CC   the European Bioinformatics Institute. There are
CC   use by non-profit institutions as long as its
CC   modified and this statement is not removed. Usage
CC   entities requires a license agreement (See http://
CC   or send an email to license@isb-sib.ch).
CC   ---------- ----------  -  ---  ------------
DR   EMBL; U67954; AAB52[illegible].1; -.
```

```
DR   WORMPEP; F41D9.5; CE10242.
DR   PFAM; PF00916; Sulfate_transp; 1.
DR   PROSITE; PS01130; SULFATE_TRANSP; 1.
KW   Hypothetical protein; Transmembrane; Transport.
FT   TRANSMEM     37     57       POTENTIAL.
FT   TRANSMEM     60     80       POTENTIAL.
FT   TRANSMEM     82    102       POTENTIAL.
FT   TRANSMEM    109    129       POTENTIAL.
FT   TRANSMEM    192    212       POTENTIAL.
FT   TRANSMEM    228    248       POTENTIAL.
FT   TRANSMEM    270    290       POTENTIAL.
FT   TRANSMEM    294    314       POTENTIAL.
FT   TRANSMEM    335    355       POTENTIAL.
FT   TRANSMEM    375    395       POTENTIAL.
FT   TRANSMEM    409    429       POTENTIAL.
FT   TRANSMEM    473    493       POTENTIAL.
SQ   SEQUENCE   758 AA;  84958 MW;  721B87729F2E12F3 CRC64;

  Query Match          14.3%;  Score 81;  DB 1;  Length 758;
  Best Local Similarity  27.6%;  Pred. No. 1.43e+00;
  Matches   16;  Conservative   14;  Mismatches  26;  Indels   2;  Gaps   2;

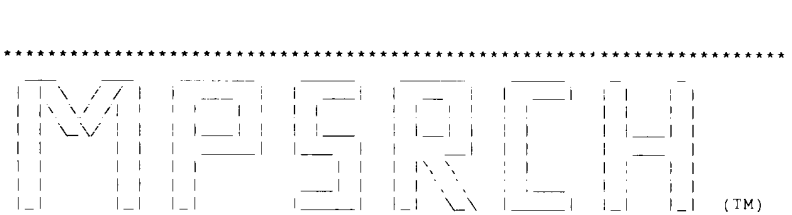
Db     306 PVPFELVLLTHYLINRFPPPSLPRTDLIRHIGLNAAAIATTAVAIHITVAKVVEKRYK 363
           || | |:  |::: |   : || | |:::: | |:  :    | |   : :   |:
Qy       3 PV-FSLLWTIQFFANARQSTSLPPFLLFKNLLLAAIETSII-ERA-HNESPSYIRHPYR 58


RESULT   15
ID   CGT_RAT         STANDARD;      PRT:   541 AA.
AC   Q09426;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   2-HYDROXYACYLSPHINGOSINE 1-BETA-GALACTOSYLTRANSFERASE PRECURSOR
DE   (EC 2.4.1.45) (UDP-GALACTOSE-CERAMIDE GALACTOSYLTRANSFERASE) (CERAMIDE
DE   UDP-GALACTOSYLTRANSFERASE) (CEREBROSIDE SYNTHASE).
GN   CGT OR UGT4.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WISTAR; TISSUE=BRAIN;
RX   MEDLINE; 94052143.
RA   Schulte S., Stoffel W.;
RT   "Ceramide UDPgalactosyltransferase from myelinating rat brain:
RT   purification, cloning, and expression.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:10265-10269(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX   MEDLINE; 94358923.
RA   Stahl N., Borvitz M., Morell P., Suzuki K., Popko B.;
RT   "Isolation, characterization, and expression of cDNA clones that
RT   encode rat UDP-galactose: ceramide galactosyltransferase.";
RL   J. Neurosci. Res. 38:234-242(1994).
CC   -!- FUNCTION: CATALYZES THE TRANSFER OF GALACTOSE TO CERAMIDE, A KEY
CC       ENZYMATIC STEP IN THE BIOSYNTHESIS OF GALACTOCEREBROSIDES, WHICH
CC       ARE ABUNDANT SPHINGOLIPIDS OF THE MYELIN MEMBRANE OF THE CENTRAL
CC       NERVOUS SYSTEM AND PERIPHERAL NERVOUS SYSTEM.
CC   -!- CATALYTIC ACTIVITY: UDP-GALACTOSE + 2-(2-HYDROXYACYL)SPHINGOSINE
CC       = UDP + 1-(BETA-D-GALACTOSYL)-2-(2-HYDROXYACYL)SPHINGOSINE.
CC   -!- PATHWAY: INVOLVED IN THE BIOSYNTHESIS OF GALACTOCEREBROSIDE.
CC   -!- TISSUE SPECIFICITY: BRAIN, RESTRICTED TO THE OLIGODENDROCYTE-
CC       CONTAINING CELL LAYERS OF CEREBRUM AND CEREBELLUM.
CC   -!- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
CC   ---------------------------  ----------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------  ----------------------------------------
DR   EMBL; L21698; AAA16108.1; -.
DR   EMBL; U07683; AAA50212.1; -.
DR   PFAM; PF00201; UDPGT; 1.
DR   PROSITE; PS00375; UDPGT; 1.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal;
KW   Microsome.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21    541       2-HYDROXYACYLSPHINGOSINE 1-BETA-
FT                                GALACTOSYLTRANSFERASE.
FT   TRANSMEM    472    492       POTENTIAL.
FT   CARBOHYD     78     78       POTENTIAL.
FT   CARBOHYD    333    333       POTENTIAL.
FT   CARBOHYD    442    442       POTENTIAL.
SQ   SEQUENCE   541 AA;  61126 MW;  260D7603170151BB CRC64;

  Query Match          14.1%;  Score 80;  DB 1;  Length 541;
  Best Local Similarity  32.6%;  Pred. No. 2.01e+00;
  Matches   15;  Conservative   14;  Mismatches  16;  Indels   1;  Gaps   1;

Db       6 PYFMLLWSAVGIAPAAKIIIVPPIMFFSHLYIFKTLASALHERGHH 51
           | | |||::  :| |     :|||::  :| :  | :||: ||:|:
Qy       3 PVFSLLWTIQFFANARQSTSLPPFLLFKNLLLAAIETSII-ERAHN 47


Search completed: Fri Aug 25 17:49:12 2000
Job time : 17 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 25 17:49:29 2000: MasPar time 9.95 Seconds
536.490 Million cell updates/sec

Tabular output not generated.

Title: >US-09-257-585-3
Description: (1-77) from US09257585.pep
Perfect Score: 568
Sequence: 1 MGPVFSLLWTTQFFANARQS.....RPCGLLSSSQCLERLHHPTIL 77

Scoring table: PAM 150
Gap 11

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl12
1:sp_archea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 38.134; Variance 63.367; scale 0.602

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	92	16.2	285	10	Q24526	MRP-LIKE ABC TRANSPORT	6.32e-02
2	92	16.2	1622	10	Q48908	MGATP ENERGIZED GLUTAT	6.32e-02
3	92	16.2	1623	10	Q64590	MULTIDRUG RESISTANCE-A	6.32e-02
4	92	16.2	1623	10	Q48997	MGATP ENERGIZED GLUTAT	6.32e-02
5	91	16.0	97	8	Q9XMR8	ORF97.	9.14e-02
6	91	16.0	1623	10	Q22449	MULTIDRUG RESISTANCE-A	9.14e-02
7	90	15.8	1136	3	P87224	CYA5 REQUIRED FOR COX1	1.32e-01
8	90	15.8	1861	5	Q01401	MICROTUBULE ASSOCIATED	1.32e-01
9	89	15.7	140	5	Q21248	COSMID F08F1.	1.90e-01
10	89	15.7	249	2	Q9X2W5	HYPOTHETICAL 28.1 KD P	1.90e-01
11	89	15.7	1622	10	Q24635	MULTIDRUG RESISTANCE-A	1.90e-01
12	88	15.5	643	5	Q02939	COSMID T23G5.	2.72e-01
13	87	15.3	221	2	Q57982	RNASE III.	3.89e-01
14	86	15.1	219	2	Q57319	HYPOTHETICAL 25.1 KD P	5.56e-01
15	84	14.8	389	10	Q24175	RFL.	1.12e+00
16	84	14.8	931	3	O60009	UBIQUITIN FUSION DEGRA	1.12e+00
17	83	14.6	155	2	Q54791	ISPO PROTEIN.	1.59e+00
18	83	14.6	241	2	P73464	LUMQ.	1.59e+00
19	83	14.6	365	2	O86574	OLIGOPEPTIDE ABC TRANS	1.59e+00
20	83	14.6	893	4	O75555	ABC TRANSPORTER MOAT-B	1.59e+00
21	83	14.6	1325	4	Q9YE32	ABC TRANSP	1.59e+00
22	82	14.4	238	1	Q17721	238AA LONG	2.24e+00
23	82	14.4	370	2	Q9ZJY9	CARBAMOYL-P	2.24e+00
24	81	14.3	256	1	Q26235	COBALT TRANS	3.15e+00
25	81	14.3	494	3	P74024	HYPOTHETICA	3.15e+00
26	81	14.3	487	2	Q87442	PUTATIVE RE	3.15e+00
27	80	14.1	372	5	Q18933	D1046.2 PRO	4.43e+00
28	80	14.1	438	10	Q25789	P4P9.7 PROT	4.43e+00
29	80	14.1	501	2	Q58545	BYE-1&EIIA	4.43e+00
30	80	14.1	638	8	Q33749	NADH DEHYDR	4.43e+00
31	80	14.1	780	10	Q40485	VOLTAGE-DEP	4.43e+00
32	80	14.1	1345	1	Q54457	STABLE PROT	4.43e+00
33	79	13.9	181	3	Q9Y702	WD REPEAT P	6.23e+00
34	79	13.9	240	5	Q225R9	T19E10.2 PR	6.23e+00
35	79	13.9	582	1	Q29749	NITRATE REI	6.23e+00
36	79	13.9	633	5	Q9XVS9	C15B11.4 PR	6.23e+00
37	79	13.9	857	14	Q12381	POL (PROVIR	6.23e+00
38	79	13.9	527	5	Q9XXI9	Y39A1A.5A P	6.23e+00
39	79	13.9	530	8	Q36016	RIBOSOMAL P	6.23e+00
40	79	13.9	750	3	Q41858	SER-LYS RIC	6.23e+00
41	79	13.9	884	7	Q29875	MHC CLASS I	6.23e+00
42	79	13.9	3014	14	Q42969	POLYPROTEIN	6.23e+00
43	78	13.7	427	5	Q94584	BEATEN PATH	8.23e+00
44	78	13.7	542	5	Q18667	C45B11.1A P	8.23e+00
45	78	13.7	731	4	Q96772	PKD2L (FRAG	8.23e+00

ALIGNMENTS

RESULT 1
ID Q24526 PRELIMINARY; PRT; 285 AA.
AC Q24526;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update
DE MRP-LIKE ABC TRANSPORTER (FRAGMENT).
OS Arabidopsis thaliana (Mouse ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyt
OC euphyllophytes, Spermatophyta, Magnoliophyta; eud.
OC core eudicots; Rosidae; eurosids II; Brassicales;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97415407.
RA TOMMASINI R., VOGT E., FROMMI E., FRAMENTAU M., AM
RA MARTINOIA E.;
RT "Differential expression of genes coding for ABC t
RT treatment of Arabidopsis thaliana with xenobiotic
RL FEBS Lett. 411:206-210(1997).
DR EMBL; U96400; AAC49708.1; -.
DR MENDEL; 24508; Arath1944.45pm.
DR PFAM; PF00005; ABC tran; 1.
FT NON_TER 1 1
SQ SEQUENCE 285 AA, 31179 MW; 559AFC88 CRC32;

Query Match 16.2%; Score 92; DB 10; Le
Best Local Similarity 26.7%; Pred. No. 6.32e-02;
Matches 16; Conservative 15; Mismatches 26;

Db 198 FSVGQRQLLSLSRALLKKSKILVLDEATAAVDVRTDALIQKTIR
Qy 12 FFANARQSTSLPPFLLRNLLAAIF

RESULT 2
ID Q48908 PRELIMINARY; PRT; 1622 AA.
AC Q48908;
DT 01-JUN-1998 (TrEMBLrel. 01, Created)
DT 01-JUN-1998 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update
DE MGATP ENERGIZED GLUTATHIONE S-C NJUGATE PUMP (FR
GN MRP2.
OS Arabidopsis thaliana (Mouse ear cress).

```
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC   Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   LU Y.-P., LI Z.-S., DROZDOWICZ Y.M., HORTENSTEINER S., MARTINOIA E.,
RA   REA P.A.;
RL   Plant Cell 10:1-18(1998).
DR   EMBL; AF020289; AAC04246.1; -.
DR   HSSP; P13569; 1NBD.
DR   MENDEL; 24661; Arath;1944;24661.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 2.
DR   PFAM; PF00664; ABC_membrane; 2.
DR   PFAM; PF00005; ABC_tran; 2.
KW   Hydrolase; ATP-binding; Transport.
SQ   SEQUENCE   1622 AA;  182042 MW;  A1A1D5D2 CRC32;

  Query Match           16.2%;  Score 92;  DB 10;  Length 1622;
  Best Local Similarity  26.7%;  Pred. No. 6.32e-02;
  Matches   16;  Conservative   16;  Mismatches  26;  Indels   2;  Gaps   2;

Db    1378 FSVGQRQLLSLSRALLRRSKILVLDEATAAVDVPTDALIQKTIREEFKSCTMLIIAHRLN 1437
           |    ||  ||:  || :: :|   | |:::: |:       ||  :::| :|  :: |:
Qy      13 FFANARQSTSLPPFLLFKNLLLAATE-TSTIE-RAHNESPSYIRHPYRPCGLLSSSQCLE 70


RESULT    3
ID   O64590      PRELIMINARY;      PRT;  1623 AA.
AC   O64590;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   MULTIDRUG RESISTANCE-ASSOCIATED PROTEIN, ATMRP2.
GN   T29F13.13 OR T3E10.30.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC   Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   ROUNSLEY S.D., LIN X., KETCHUM K.A., CROSBY M.I., BRANDON R.C.,
RA   SYKES S.M., KAUL S., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,
RA   SOMERVILLE C.R., VENTER J.C.;
RL   Submitted (MAY 1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC003096; AAC16268.1; -.
DR   EMBL; AC004077; AAC26713.1; -.
DR   HSSP; P13569; 1NBD.
DR   MENDEL; 29399; Arath;1944;29399.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 2.
DR   PFAM; PF00664; ABC_membrane; 2.
DR   PFAM; PF00005; ABC_tran; 2.
KW   ATP-binding; Transport.
SQ   SEQUENCE   1623 AA;  182129 MW;  525D4A5B CRC32;

  Query Match           16.2%;  Score 92;  DB 10;  Length 1623;
  Best Local Similarity  26.7%;  Pred. No. 6.32e-02;
  Matches   16;  Conservative   16;  Mismatches  26;  Indels   2;  Gaps   2;

Db    1379 FSVGQRQLLSLSRALLRRSKILVLDEATAAVDVPTDALIQKTIREEFKSCTMLIIAHRLN 1438
           |    ||  ||:  || :: :|   | |:::: |:       ||  :::| :|  :: |:
Qy      13 FFANARQSTSLPPFLLFKNLLLAATE-TSTIE-RAHNESPSYIRHPYRPCGLLSSSQCLE 70


RESULT    4
ID   O48907      PRELIMINARY;      PRT;  1623 AA.
AC   O48907;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   MGATP-ENERGIZED GLUTATHIONE S-CONJUGATE PUMP (EC 3.6.1.3).
GN   MRP2.
OS   Arabidopsis thaliana (Mouse ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC   Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   LU Y.-P., LI Z. S., DROZDOWICZ Y.M., HORTENSTEINER S., MARTINOIA E.,
RA   REA P.A.;
RL   Plant Cell 10:1-18(1998).
DR   EMBL; AF020288; AAC04245.1; -.
DR   HSSP; P13569; 1NBD.
DR   MENDEL; 24660; Arath;1944;24660.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 2.
DR   PFAM; PF00664; ABC_membrane; 2.
DR   PFAM; PF00005; ABC_tran; 2.
KW   Hydrolase; ATP-binding; Transport.
SQ   SEQUENCE   1623 AA;  182072 MW;  ABD1DF16 CRC32;

  Query Match           16.2%;  Score 92;  DB 10;  Length 1623;
  Best Local Similarity  26.7%;  Pred. No. 6.32e-02;
  Matches   16;  Conservative   16;  Mismatches  26;  Indels   2;  Gaps   2;

Db    1379 FSVGQRQLLSLSRALLRRSKILVLDEATAAVDVPTDALIQKTIREEFKSCTMLIIAHRLN 1438
           |    ||  ||:  || :: :|   | |:::: |:       ||  :::| :|  :: |:
Qy      13 FFANARQSTSLPPFLLFKNLLLAATE-TSTIE-RAHNESPSYIRHPYRPCGLLSSSQCLE 70


RESULT    5
ID   Q9XMR8      PRELIMINARY;      PRT;    97 AA.
AC   Q9XMR8;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   ORF97.
OS   Tetrahymena pyriformis.
OG   Mitochondrion.
OC   Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC   Tetrahymenina; Tetrahymena.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   EDQVIST J., BURGER G., GRAY M.W.;
RT   "Expression of mitochondrial protein coding genes in Tetrahymena
RT   pyriformis.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   BURGER G., ZHU Y., LITTLEJOHN T.G., GREENWOOD S.J., SCHNARE M.N.,
RA   LANG B.F., GRAY M.W.;
RT   "Complete sequence, gene content and organization of the mitochondrial
RT   genome of Tetrahymena pyriformis. Comparison with Paramecium aurelia
RT   mitochondrial DNA.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF160864; AAD41946.1; -.
KW   Mitochondrion.
SQ   SEQUENCE   97 AA;  11991 MW;  00224B5B CRC32;

  Query Match           16.0%;  Score 91;  DB 8;  Length 97;
  Best Local Similarity  41.4%;  Pred. No. 9.14e-02;
  Matches   12;  Conservative    7;  Mismatches   8;  Indels   2;  Gaps   2;

Db      49 STVVELIWPTHFF-NRRHTSS-QEFLKFR 65
           : |  |:| |:|| | |:::|   || |:
Qy       2 GPVFSLLWTTQFFANARQSTSLPPFLLFK 30


RESULT    6
ID   O22449      PRELIMINARY;      PRT;  1623 AA.
AC   O22449;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
```

```
DE   MULTIDRUG RESISTANCE-ASSOCIATED PROTEIN 2.
GN   MRP2.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta, Tracheophyta,
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC   Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LANDSBERG ERECTA;
RA   MARIN E., PRAT N., VAVASSEUR A., FORESTIER C.;
RL   Biochim. Biophys. Acta 0:0-0(1997).
DR   EMBL; AF014960; AAC49988.1; -.
DR   HSSP; P14569; 1NBD.
DR   MENDEL; 24317; Arath;1944;24417.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 2.
DR   PFAM; PF00664; ABC_membrane; 2.
DR   PFAM; PF00005; ABC_tran; 2.
KW   ATP-binding; Transport.
SQ   SEQUENCE   1623 AA;  181967 MW;  E8B07BA2 CRC32;

  Query Match             16.0%;  Score 91;  DB 10;  Length 1623;
  Best Local Similarity  26.7%;  Pred. No. 9.14e-02;
  Matches   16;  Conservative   16;  Mismatches  26;  Indels   2;  Gaps   2;

Db   1379 FSVGQRQLLSLSRCLLRRSKILVLDEATAAVDVRTDALIQKTIREEFKSCTMLIIAHRLN 1438
           |   ||  ||:  || :: :|   | |:::: |:        ||  :::| :|  :: |:
Qy     13 FFANARQSTSLPPFLLFKNLLAATETSTIERAHN 70


RESULT    7
ID   P87224       PRELIMINARY;      PRT;  1136 AA.
AC   P87224;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   CYA5 REQUIRED FOR COX1 TRANSLATION.
GN   CYA5.
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Euascomycetes; Pyrenomycetes;
OC   Sordariales; Sordariaceae; Neurospora.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=74A;
RA   COFFIN J.W., DHILLON R., RITZEL R.G., NARGANG F.E.;
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF002169; AAB60884.1; -.
DR   PFAM; PF01535; DUF17; 1.
SQ   SEQUENCE   1136 AA;  129918 MW;  BAAB95C9 CRC32;

  Query Match             15.8%;  Score 90;  DB 3;  Length 1136;
  Best Local Similarity  35.0%;  Pred. No. 1.32e-01;
  Matches   14;  Conservative   14;  Mismatches  10;  Indels   2;  Gaps   2;

Db    847 LYTIQF-PSAKEPSNLPPRLL NNMMLADYHERKFDRVRE 884
          | |||| ::|:::::::|  || :|:::||      :|| ::
Qy      8 LWTIQFFANARQSTSLPPFLLFKNLLAATETSTIERAHN 47


RESULT    8
ID   O01401       PRELIMINARY;      PRT;  1861 AA.
AC   O01401;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   MICROTUBULE ASSOCIATED PROTEIN.
GN   ASP.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON R;
RA   SAUNDERS R.D.C., AVILES M.V., EDWARD T.I.A., O'HARE K.,
RA   GLOVER D.M.G.;
RL   J. Cell Biol. 0:0-0(0).
DR   EMBL; U95171; AAB51540.1; -.
DR   FLYBASE; FBgn0000140; asp.
DR   PFAM; PF00612; IQ; 12.
SQ   SEQUENCE   1861 AA;  219598 MW;  12FDF4AF CRC32;

  Query Match             15.8%;  Score 90;  DB 5;  Length
  Best Local Similarity  44.6%;  Pred. No. 1.32e-01;
  Matches   18;  Conservative   5;  Mismatches  10;

Db   1003 LFAKKLLAAAETAKQKSQKQLAAASYQMQWRTYQL-GPTQRHI
          || : :|||:|: ::|:::::  ::: :  :  : :    :
Qy     28 LFKNLLLAATETSTIERAHN  JFSYTRHFYRPGGLLSSSQCLE


RESULT    9
ID   Q21248       PRELIMINARY;      PRT;   140 AA.
AC   Q21248;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   COSMID K05F1.
GN   K05F1.10.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdi
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., B
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., Doe
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LAIP
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B.,
RA   PARSONS J., PERCY C., RIFKEN L., ROPELEWSKA E., SANDER
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SU
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHL
RT   "2.2 Mb of contiguous nucleotide sequence from chr
RT   elegans.";
RL   Nature 368:32 38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   WOHLDMANN P.;
RL   Submitted (JUN-1995) to the EMBL/GenBank/DDBJ data
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   WATERSTON R.;
RL   Submitted (JUN-1995) to the EMBL/GenBank/DDBJ data
DR   EMBL; U29377; AAA68717.1;
DR   HSSP; P19398; 1ATD.
SQ   SEQUENCE   140 AA;  16371 MW;  E2D88621 CRC32;

  Query Match             15.7%;  Score 89;  DB 5;  Length
  Best Local Similarity  47.8%;  Pred. No. 1.90e-01;
  Matches   11;  Conservative   7;  Mismatches  4;

Db     66 RHCDRTCENLFSPPHCLNHLHHA 88
          ||  | |  |:::::|:::|:
Qy     54 RHPYRPC-GLLSSSQCLERLHHA 75


RESULT   10
ID   Q9X2W5       PRELIMINARY;      PRT;   249 AA.
AC   Q9X2W5;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
```

```
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   HYPOTHETICAL 28.0 KD PROTEIN.
OS   Butyrivibrio fibrisolvens.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Butyrivibrio.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 99240441.
RA   KALMOKOFF M.L., LU D., WHITFORD M.F., TEATHER R.M.;
RT   "Evidence for production of a new lantibiotic (Butyrivibriocin OR79A)
RT   by the ruminal anaerobe butyrivibrio fibrisolvens OR79:
RT   characterization of the structural gene encoding butyrivibriocin
RT   OR79A.";
RL   Appl. Environ. Microbiol. 65:2128-2135(1999).
DR   EMBL; AF062647; AAD24813.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   249 AA;  27960 MW;  EB30C041 CRC32;

  Query Match           15.7%;  Score 89;  DB 2;  Length 249;
  Best Local Similarity  42.5%;  Pred. No. 1.98e-01;
  Matches   13;  Conservative   12;  Mismatches  14;  Indels   1;  Gaps   1;

Db    151 MIPAVALMWAITVIFEKPILSVGLNIFLVLPGILVANTGA 190
          | |  :|:|: | :| ::  | :|  ||::  :|:| | :
Qy      1 MGPVFSLLWT-TQFFANAPQSTSLPPFLLFKNLLLAATET 39


RESULT   11
ID   O24635      PRELIMINARY;      PRT;  1622 AA.
AC   O24635;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   MULTIDRUG RESISTANCE-ASSOCIATED PROTEIN HOMOLOG.
GN   MRP1.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC   Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97368351.
RA   LU Y.P., LI Z.S., REA P.A.;
RT   "AtMRP1 gene of Arabidopsis encodes a glutathione S-conjugate pump:
RT   isolation and functional definition of a plant ATP-binding cassette
RT   transporter gene.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:8243-8248(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   LU Y.-P., LI Z.-S., REA P.A.;
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   LU Y.-P., LI Z.-S., REA P.A.;
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF008125; AAB71832.1; -.
DR   EMBL; AF008124; AAB67319.1; -.
DR   HSSP; P13569; 1NBD.
DR   MENDEL; 24381; Arath;1944;24381.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 2.
DR   PFAM; PF00664; ABC_membrane; 2.
DR   PFAM; PF00005; ABC_tran; 2.
KW   Hydrolase; ATP-binding; Transport.
SQ   SEQUENCE   1622 AA;  182015 MW;  95065CDF CRC32;

  Query Match           15.7%;  Score 89;  DB 10;  Length 1622;
  Best Local Similarity  26.7%;  Pred. No. 1.90e-01;
  Matches   16;  Conservative   16;  Mismatches  26;  Indels   2;  Gaps   2;

Db   1474 PSVGQPQIISIARALIRRSKIIVLDEATAAVDVRTDVLIQKTIREEFKSCTMLIIAHRLN 1433
          |    ||  ||:  || :: :|   | |:::: |:       ||  :::| :|  :: |:
Qy     13 FFANARQSTSLPPFLLFKNLLLAATETSTIERAHNESPSYIEHDYRPCGLISSSQCLE 70


RESULT   12
ID   Q02039      PRELIMINARY;      PRT;  533 AA.
AC   Q02039;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   COSMID T23C6.
GN   T23C6.3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAK S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALEN N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   DU Z.;
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   WATERSTON R.;
RL   Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF000191; AAB52884.1; -.
SQ   SEQUENCE   533 AA;  61479 MW;  308B7149 CRC32;

  Query Match           15.5%;  Score 88;  DB 5;  Length 533;
  Best Local Similarity  23.8%;  Pred. No. 2.72e-01;
  Matches   10;  Conservative   17;  Mismatches  14;  Indels   1;  Gaps   1;

Db    472 FIPTIPSTSTMPDMMPYPNVLLTAA PGDVAPIDSLTPSFLR 512
          |::    ::::|| : : |:||:|:  : :    : :|||:|
Qy     13 FFANARQSTSLPPFLLFKNLLLAATETSTIERAHNESPSYIR 54


RESULT   13
ID   O67082      PRELIMINARY;      PRT;  221 AA.
AC   O67082;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   RNASE III.
GN   RNC.
OS   Aquifex aeolicus.
OC   Bacteria; Aquificales; Aquificaceae; Aquifex.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VF5;
RX   MEDLINE; 98196666.
RA   DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
RA   GRAHAM D.E., OVERBEEK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
RA   FELDMAN R.A., SHORT J.M., OLSEN G.J., SWANSON R.V.;
RT   "The complete genome of the hyperthermophilic bacterium Aquifex
RT   aeolicus.";
RL   Nature 392:353-358(1998).
```

```
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN-VF5;
RA   DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
RA   GRAHAM D.E., OVERBEEK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
RA   FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RL   Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE000715; AAC07049.1; -.
DR   PFAM; PF00035; dsrm; 1.
DR   PFAM; PF00636; Ribonuclease_3; 1.
SQ   SEQUENCE   221 AA;  26109 MW;  53BA1091 CRC32;

  Query Match           15.3%;  Score 87;  DB 2;  Length 221;
  Best Local Similarity  24.5%;  Pred. No. 2.89e-01;
  Matches   12;  Conservative   17;  Mismatches  18;  Indels   2;  Gaps   2;

Db    105 IGDVFEALWAAVYIDSGRDANFTRFIFYFLEKFDILSAIKEGPVKEDYK 153
          :| ||  ||:: :: ::|::  :   |  |||: :|:|   : : :  :
Qy      1 MGDVFSLLWLTQFFANARQSTSLPPFLLFKNLLLAATETSIIERAHN 47


RESULT   14
ID   O67319       PRELIMINARY;      PRT;   219 AA.
AC   O67319;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   HYPOTHETICAL 25.0 KD PROTEIN.
GN   AQ_1287.
OS   Aquifex aeolicus.
OC   Bacteria; Aquificales; Aquificaceae; Aquifex.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN-VF5;
RX   MEDLINE; 98196666.
RA   DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
RA   GRAHAM D.E., OVERBEEK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
RA   FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RT   "The complete genome of the hyperthermophilic bacterium Aquifex
RT   aeolicus.";
RL   Nature 392:353-358(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN-VF5;
RA   DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
RA   GRAHAM D.E., OVERBEEK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
RA   FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RL   Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE000732; AAC07280.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   219 AA;  25050 MW;  0A339273 CRC32;

  Query Match           15.1%;  Score 86;  DB 2;  Length 219;
  Best Local Similarity  43.8%;  Pred. No. 5.56e-01;
  Matches   14;  Conservative    6;  Mismatches  10;  Indels   2;  Gaps   2;

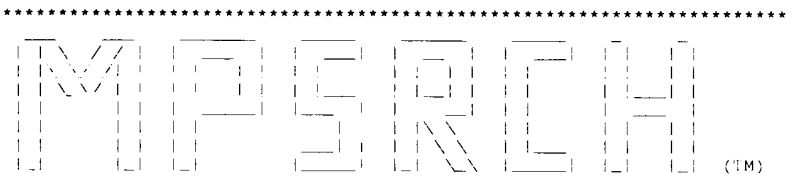
Db     13 VFGLFLFSLIFFGLLSLAT-FPKFLLFDRLLI 43
          ||:| | :  ||:    :| :| ||||  ||:
Qy      4 VFSL-LWTTQFFANARQSTSLPPFLLFKNLLL 34


RESULT   15
ID   O24175       PRELIMINARY;      PRT;   389 AA.
AC   O24175;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   RFL.
GN   RFL.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
OC   Poaceae; Oryza.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN CV. TORIDE; TISSUE LEVEL: PING PANICLE;
RA   KYOZUKA J., KONISHI S., NEMOTO K., IZAWA T., SHIMAM
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ datab
DR   EMBL; AB005620; BAA21547.1; -.
DR   MENDEL; 25649; Oryza;14291;2.48.
SQ   SEQUENCE   389 AA;  42550 MW;  03509BA0 CRC32;

  Query Match           14.8%;  Score 84;  DB 10;  Len
  Best Local Similarity  41.7%;  Pred. No. 1.12e+00;
  Matches   13;  Conservative    9;  Mismatches  19;

Db    225 PVVTEPGIVAMAKENL..PTIN...P.PIELLVLSMASIR 26
          :  || : : |: :::  :          | |   | : ::
Qy     34 LAATETSIIERAHNESLSYIPHTPP.IL-LSSSQCLERLR 74


Search completed: Fri Aug 25 17:50:52 2000
Job time : 83 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 25 17:43:45 2000; MasPar time 3.18 Seconds
335.185 Million cell updates/sec

Tabular output not generated.

Title: >US-09-257-585-2
Description: (1-45) from US09257585.pep
Perfect Score: 345
Sequence: 1 MTNKYPVKSICCSIIFNEASIYCQGKFINGISVLPELDNSVICKR 45

Scoring table: PAM 150
Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq36
1:geneseqp

Statistics: Mean 24.053; Variance 86.439, scale 0.278

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	78	23.3	365	1	W18289	Mangifera indica ACC s	9.38e-00
2	74	22.1	769	1	W04197	Env gene of simian imm	2.15e-01
3	73	21.9	494	1	W00862	Porcine E selectin.	2.64e-01
4	72	21.5	420	1	W59282	Canine herpes virus OR	3.24e-01
5	72	21.5	730	1	P82881	Lupin exo-(1-4)beta-D-	3.24e-01
6	71	21.2	530	1	W26428	Swinepox virus HindIII	3.57e-01
7	70	20.9	120	1	P42999	Human E selectin lecti	4.86e-01
8	70	20.9	120	1	R43001	Human E-selectin lecti	4.86e-01
9	70	20.9	359	1	R72715	hisC gene product of S	4.86e-01
10	70	20.9	610	1	W46733	Amino acid sequence of	4.86e-01
11	70	20.9	610	1	R05538	Endothelial-leukocyte	4.86e-01
12	70	20.9	610	1	R05494	Endothelial leukocyte	4.86e-01
13	70	20.9	610	1	R08116	Endothelial cell-leuco	4.86e-01
14	70	20.9	610	1	W18839	E-selectin.	4.86e-01
15	69	20.6	452	1	R89280	CM-cellulase.	5.95e-01
16	68	20.3	443	1	W21014	H. pylori cytoplasmic	7.27e-01
17	67	20.0	206	1	R66606	Flower style-specific	8.87e-01
18	67	20.0	206	1	W14914	Tomato S-ribonuclease	8.87e-01
19	66	19.7	646	1	W20657	H. pylori cytoplasmic	1.08e+02
20	66	19.7	943	1	R12802	170kD heavy chain of t	1.08e+02
21	66	19.7	1291	1	R68975	Entamoeba histolytica	1.08e+02
22	66	19.7	1292	1	R70681	Entamoeba histolytica	1.08e+02
23	65	19.4	102	1	R46087	Zinc resistant like pr	1.32e+02
24	65	19.4	390	1	W09879	ACC synthase	1.32e+02
25	65	19.4	495	1	W00258	Pelargonium	1.32e+02
26	64	19.1	170	1	R99275	Chemokine r	1.60e+02
27	64	19.1	537	1	R72814	Sequence of	1.60e+02
28	64	19.1	546	1	R98898	Nuclear hor	1.60e+02
29	64	19.1	654	1	W14254	Z. japonica	1.60e+02
30	64	19.1	906	1	W48795	Novel human	1.60e+02
31	64	19.1	1296	1	W08449	Caenorhabdi	1.60e+02
32	63	18.8	188	1	R11149	Ser-Gln15	1.93e+02
33	63	18.8	285	1	W08456	H. pylori C	1.93e+02
34	63	18.8	279	1	R94529	Korean Vip	1.93e+02
35	63	18.8	334	1	R45741	Myoinositol	1.93e+02
36	63	18.8	1559	1	W13905	Human SMCY	1.93e+02
37	62	18.5	168	1	W73233	Porcine in	2.29e+02
38	62	18.5	284	1	W08911	Elicitor r	2.29e+02
39	62	18.5	500	1	W26419	Swinepox vi	2.29e+02
40	62	18.5	647	1	W20069	Glucan elic	2.29e+02
41	62	18.5	667	1	R88612	Glucan elic	2.29e+02
42	62	18.5	681	1	W08915	IGER protei	2.29e+02
43	62	18.5	877	1	W08914	IER protein	2.29e+02
44	62	18.5	1054	1	W08918	CER protein	2.29e+02
45	62	18.5	1087	1	W90121	Human hnRNP	2.29e+02

ALIGNMENTS

RESULT 1
ID W18289 standard; Protein; 365 AA.
AC W18289;
DT 27-JAN-1998 (first entry)
DE Mangifera indica ACC synthase miacc1.
KW ACC synthase; S-adenosyl-L methionine methylthioa
KW miacc1; ethylene biosynthesis; transgenic plant;
KW antisense expression system; plant development; f
KW EC 4.4.1.14; mango; multigene family.
OS Mangifera indica.
FH Key Location/Qualifiers
FT Protein 1..365
FT /label= miacc1
FT /note= "Sequence represents 75% of
FT ACC synthase protein (lac
FT or stop codons)."
PN WO9711166-A1.
PD 27-MAR-1997.
PF 20-SEP-1996; AU0591.
PR 02-MAY-1996; AU-009603.
PR 20-SEP-1995; AU-005595.
PA (UYQU) UNIV QUEENSLAND.
PI Botella JR;
DR WPI; 97-202875/18.
DR N-PSDB; R72631.
PT Pineapple, papaya and mango ACC synthase genes -
PT therapy to produce fruits with reduced senescence
FS Disclosure; Fig 4; 46pp; English.
CC This sequence represents a novel enzyme, miacc1, A
CC synthase multigene family found in mango. ACC syn
CC the pathway for ethylene biosynthesis and the rat
CC expression of ACC synthase is considered to limit
CC of ethylene production. Endogenous ethylene is of
CC crops, especially if some form of mechanical wou
CC diminishes their post harvest quality and storage A
CC synthase genes expressed in transgenic plants usi
CC antisense expression systems may be used to contro
CC plant development, in particular fruit ripening,
CC thus improving storage life.
SQ Sequence 365 AA;

Query Match 23.3%; Score 78; DB 1; Len
Best Local Similarity 29.6%; Pred. No. 9.38e-00;
Matches 8; Conservative 9; Mismatches 9;

Db 188 CDEIYA AILEGFEAPVSVAEVIQEMD 213

```
Qy      12 CSIIFNEASIYCQGKFINGTSVLPPLD 38


RESULT    2
ID   R04197 standard; protein; 769 AA
AC   R04197;
DT   12-JUN-1990  (first entry)
DE   Env gene of simian immunodeficiency virus.
KW   SIV; simian immunodeficiency virus; AIDS; HIV; vaccine; ss.
OS   SIV.
PN   J01289485-A.
PD   21-NOV-1989.
PF   16-JUN-1988; 119023.
PR   16-JUN-1988; JP-119023
PA   (TOFU) Toa Nenryo Kogyo KK.
DR   WPI; 90-005176/01
DR   N-PSDB; Q02829.
PT   DNA complementary to RNA of simian immuno-deficiency virus -
PT   used in vaccines and for diagnosis of AIDS
PS   Disclosure; Fig 4; 6pp; Japanese.
CC   Env gene derived from RNA of SIV from which vector plasmid pSAH121
CC   may be constructed and incorporated into an E.coli vector. Useful in
CC   developing a vaccine against and in diagnosis of AIDS.
SQ   Sequence   769 AA;

  Query Match             22.1%;  Score 74;  DB 1;  Length 769;
  Best Local Similarity  44.0%;  Pred. No. 2.15e+01;
  Matches   11;  Conservative    4;  Mismatches   6;  Indels   1;  Gaps   1;

Db    178 KKNYYSVVWNDAEIYCKNK-INSTS 201
          |    ||: |:| ||| |  |:||
Qy      8 KSTCCSIIFNEASIYCQGKFINGTS 32


RESULT    3
ID   W99892 standard; Protein; 484 AA.
AC   W99892;
DT   10-JUN-1999  (first entry)
DE   Porcine E-selectin.
KW   Porcine; E-selectin; diagnosis; transplant rejection, xenotransplant,
KW   cytokine; donor organ endothelium; inflammatory cell; graft tissue;
KW   acute cellular allograft rejection; cell adhesion.
OS   Sus scrofa.
PN   US5891645-A.
PD   06-APR-1999.
PF   01-JUN-1994; 252493.
PR   01-JUN-1994; US-252493
PA   (ALEX-) ALEXION PHARM INC.
PI   Evans MJ, Matis LA, Rollins S, Rother RP;
DR   WPI; 99-253853/21.
DR   N-PSDB; X19809.
PI   Nucleic acids encoding the porcine E-selectin protein - useful in
PI   diagnosing and treating rejection of xenotransplanted pig organs
PS   Claim 1; Column 47-50; 30pp; English.
CC   The present sequence is porcine E-selectin (E-sel). The nucleotide
CC   sequence encoding E-sel may be used in the diagnosis, prevention and
CC   treatment of tissue rejection following xenotransplants of pig organs
CC   and also in the monitoring (by specific measurement of the amount of
CC   E-sel in the blood of the patient) and modulation of immune responses to
CC   the transplants. Cytokine-induced expression of E sel by donor organ
CC   endothelium, contributes to the binding and subsequent transmigration of
CC   inflammatory cells into the graft tissue that occurs during acute
CC   cellular allograft rejection. Therefore, preventing expression of E sel
CC   on the donor tissue may reduce the risk of rejection by modulating E-sel
CC   mediated cell adhesion. Fragments of the nucleotide sequence encoding
CC   E-sel may be used as polymerase chain reaction (PCR) primers or
CC   hybridization probes for the identification and/or isolation of the
CC   porcine E-sel gene from genomic DNA. Antisense sequences may be used to
CC   inhibit the expression of E-sel. DNA constructs may be used to introduce
CC   antisense sequences to a cell, or to inactivate the E-sel gene (e.g.
CC   knock-out constructs). Cells in which the expression of E-sel has been
CC   prevented may be used as disease models or to produce transgenic animals
CC   (e.g. pigs) from which organs may be harvested for transplantation with
CC   a reduced risk of rejection.
SQ   Sequence   484 AA;

  Query Match             21.8%;  Score 73;  DB 1;  Length 484;
  Best Local Similarity  28.6%;  Pred. No. 2.64e+01;
  Matches    6;  Conservative    9;  Mismatches   6;  Indels   0;  Gaps   0;

Db     31 TMTFDDASAYCQQRYTHLVAI 51
          :: |::|| ||| :: :  ::
Qy     13 SIIFNEASIYCQGKFINGTSV 33


RESULT    4
ID   W50282 standard; Protein; 420 AA.
AC   W50282;
DT   29-SEP-1998  (first entry)
DE   Canine herpes virus ORF2 protein.
KW   Canine herpes virus; open reading frame, ORF, viral replication; antigen;
KW   viral vector, thymidine kinase, vaccine, infection, canine pathogen.
OS   Canine herpes virus.
PN   WO9749825-A1.
PD   31-DEC-1997.
PF   23-JUN-1997; F01115.
PR   27-JUN-1996; FR-008242.
PA   (MERI-) MERIAL.
PI   Audonnet JC, Baudu P;
DR   WPI; 98-077186/07.
DR   N-PSDB; V06946.
PT   Live recombinant vaccine based on canine herpes virus - containing
PT   at least one heterologous sequence encoding antigen, particularly
PT   from distemper virus, for use in dogs
PS   Example 2; Fig 1; 53pp; French.
CC   This sequence represents the protein encoded by the open reading frame 2
CC   (ORF2) present in the ScaI-XhoI fragment of the canine herpes virus (CHV)
CC   genome.  In this region 5 ORFs were found, of which ORFs 3 and 5 were
CC   shown not to be essential for viral replication.  The invention relates
CC   to the construction of novel vectors based on the CHV sequence, in which
CC   heterologous gene sequences can be inserted into non-essential sites such
CC   as ORF3, ORF5, the thymidine kinase gene or the region between ORF19 and
CC   ORF 22 of CHV.  Vaccines which express heterologous proteins, especially
CC   antigens of canine pathogens, are used to protect against these pathogens
CC   and simultaneously against CHV infection.
SQ   Sequence   420 AA;

  Query Match             21.5%;  Score 72;  DB 1;  Length 420;
  Best Local Similarity  54.5%;  Pred. No. 3.24e+01;
  Matches   12;  Conservative    4;  Mismatches   5;  Indels   1;  Gaps   1;

Db     87 IINNEENINSEEGKFISGRAVL 108
          || || :|  : ||||:| :||
Qy     14 IIFNEASIYCQ-GKFINGTSVL 34


RESULT    5
ID   R82881 standard; Protein; 730 AA.
AC   R82881;
DT   10-FEB-1996  (first entry)
DE   Lupin exo-(1-4)beta-D-galactanase enzyme.
KW   Exo-(1-4) beta D galactanase, enzyme, transgenic plant;
KW   crop improvement.
OS   Lupus angustifolius.
PN   WO9523228-A1.
PD   31-AUG-1995.
PF   23-FEB-1995; C00372.
PR   23-FEB-1994; GB-003423.
PA   (UNIL ) UNILEVER NV.
PA   (UNIL ) UNILEVER PLC.
PI   Chengappa S,  De SILVA J,  Hellyer SA,  Reid JSG;
DR   WPI; 95-311537/40.
DR   N-PSDB; T01014.
PT   Lupin and tomato exo-galactanase DNA and protein - useful for the
PT   alteration of plant characteristic(s), e.g. texture, growth,
PT   ripening
```

```
PS   Claim 1; Page 25; 67pp; English.
CC   This enzyme may be expressed in a transgenic plant in order to
CC   alter characterstics of e.g. growth, texture or ripening of the
CC   plant or plant parts.  This protein may also be expressed in a
CC   recombinant host for the production of the enzyme which may be used
CC   for the modification, degradation or liquefaction of plant materials
CC   in order to affect mechanical properties relating to eating texture,
CC   particle sizes of, e.g. fruit or vegetable juices, or extractability
CC   of colors, flavors or vitamins.
SQ   Sequence   730 AA;

  Query Match             21.5%;  Score 72;  DB 1;  Length 730;
  Best Local Similarity   38.5%;  Pred. No. 3.24e+01;
  Matches    15;  Conservative    6;  Mismatches   17;  Indels    1;  Gaps    1;

Db     378 YKIFSACAAAFIANYNIDYSIQVKPCNCQYDLPPWSISIL 417
           |   |:| ::: |  : |  | || ||   |||   |:|
Qy       5 YPVKSTCCSIIFNEASIYC-QGKPINGTSVLPPIDNSVL 42


RESULT    6
ID   W26428 standard; Protein; 530 AA.
AC   W26428;
DT   16-DEC-1997  (first entry)
DE   Swinepox virus HindIII C encoded protein C18L.
KW   SPV; vaccine; vector; pseudorabies virus; C18L.
OS   Swinepox virus.
PN   US5651972-A.
PD   29-JUL-1997.
PF   21-APR-1989; 342212.
PR   01-JUL-1992; US-908241.
PR   21-APR-1989; US-342212.
PR   29-JUN-1992; US-908640.
PR   14-SEP-1994; US-307499.
PA   (UYFL ) UNIV FLORIDA RES FOUND INC.
PI   Gibbs EPJ,  Moyer RW,  Vinuela E;
DR   WPI; 97-392897/36.
DR   N-PSDB; T84564.
PT   Recombinant swinepox virus vector - used particularly for vaccines
PT   against infectious agents, including pseudorabies
PS   Example 2; Column 97-101; 70pp; English.
CC   This sequence comprises a polypeptide, designated C18L, encoded
CC   by the antisense strand (open reading frame 3) of the HindIII C
CC   fragment (see T84564) of swinepox virus (SPV).  C18L shows no
CC   apparent homology to other proteins.  A claimed recombinant vector
CC   comprises a heterologous nucleotide sequence inserted into, or
CC   replacing, all or a portion of a non-essential SPV gene or nucleic
CC   acid sequence of the HindIII C fragment.  The vector can be used
CC   for the expression of heterologous proteins, both in vivo as a
CC   vaccine, and in vitro for production of the selected protein.  The
CC   heterologous protein is preferably pseudorabies virus gp50 or gp63
CC   for use in swine vaccines.  As SPV is host-restricted to swine, the
CC   use of modified recombinant SPV as a live vaccine vector eliminates
CC   the risk of spreading infection with the virus to other animal
CC   populations.
SQ   Sequence   530 AA;

  Query Match             21.2%;  Score 71;  DB 1;  Length 530;
  Best Local Similarity   37.1%;  Pred. No. 3.97e+01;
  Matches    14;  Conservative    7;  Mismatches   12;  Indels    3;  Gaps    3;

Db     415 MKLKQFPIAKSGISSIVYND-FIYCIGGRIDTPHI 448
           |  | :|: :|   ||:|:|  ||| |  |:   :
Qy       1 MINK-YPV-KSTCCSIIFNEASIYCQGKPINGTSV 33


RESULT    7
ID   R42999 standard; protein; 120 AA.
AC   R42999;
DT   23-MAY-1994  (first entry)
DE   Human E-selectin lectin domain E8A variant.
KW   Cell adhesion molecule; selectin; lectin domain; inflammation;
KW   PCR mutagenesis; E-selectin mutein; enhanced binding affinity.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   domain          1..120
FT                   /label= lectin_domain
FT                   /note= "wild-type Gln8 residue sub
PN   WO9322436-A.
PD   11-NOV-1993.
PF   28-APR-1993; U03987.
PR   30-APR-1992; US-879036.
PR   01-OCT-1992; US-956702.
PA   (GETH ) GENENTECH INC.
PA   (HOFF ) HOFFMANN LA ROCHE & CO AG F.
PI   Burns DK,  Erbe DV,  Lasky LA,  Presta LG,  Wolit
DR   WPI; 93-368795/46.
PT   New selectin mutants with modified lectin domains
PT   increased affinity for natural ligand, useful for
PT   inflammation, also related DNA, vectors, antibodie
PS   Claim 24; Fig 2A and Page 66; 91pp; English.
CC   The human E-selectin variant E8A is a preferred se
CC   sequence of the wild-type lectin domain, but not
CC   R42999, is given in the specification. The substi
CC   8 of the lectin domain results in a mutant with m
CC   affinity than wild-type E-selectin.
SQ   Sequence   120 AA;

  Query Match             21.2%;  Score 70;  DB 1;  Leng
  Best Local Similarity   28.6%;  Pred. No. 4.86e+01;
  Matches     6;  Conservative     9;  Mismatches    5;

Db       9 AMTYDEASAYCGGKYTHLVAI 29
           :: ::||| || |:: :  ::
Qy      15 SIIFNEASIYCQGKPINGTSV 35


RESULT    8
ID   R43001 standard; protein; 120 AA.
AC   R43001;
DT   23-MAY-1994  (first entry)
DE   Human E selectin lectin domain R84A, K86A variant
KW   Cell adhesion molecule; selectin; lectin domain;
KW   PCR mutagenesis, E selectin mutein, enhanced bind
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   domain          1..120
FT                   /label= lectin_domain
FT                   /note= "wild-type Arg84 and Lys86
FT                   both substit by Ala"
PN   WO9322436-A.
PD   11-NOV-1993.
PF   28-APR-1993; U03987.
PR   30-APR-1992; US-879...
PR   01-OCT-1992; US-956702.
PA   (GETH ) GENENTECH INC.
PA   (HOFF ) HOFFMANN LA ROCHE & CO AG F.
PI   Burns DK,  Erbe DV,  Lasky LA,  Presta LG,  Wolit
DR   WPI; 93-368795/46.
PT   New selectin mutants with modified lectin domains
PT   increased affinity for natural ligand, useful for
PT   inflammation, also related DNA, vectors, antibodi
PS   Claim 24; Fig 2A and Page 66; 91pp; English.
CC   The human E-selectin variant (R84A, K86A) is a pr
CC   mutant. The sequence of the wild-type lectin doma
CC   variant sequence R43001, is given in the specific
CC   substitutions at positions 84 and 86 of the lectin
CC   in a mutant with much higher binding affinity tha
CC   E-selectin.
SQ   Sequence   120 AA;

  Query Match             21.2%;  Score 70;  DB 1;  Le
  Best Local Similarity   28.6%;  Pred. No. 4.86e+01;
  Matches     6;  Conservative     9;  Mismatches    5;

Db       9 AMTYDEASAYCGGKYTHLVAI 29
```

```
          :: ::||| || |:: :  ::
Qy      13 SIIFNEASIYCQGKFINGTSV 33


RESULT    9
ID   R72715 standard; Protein; 459 AA.
AC   R72715;
DT   02-DEC-1995  (first entry)
DE   hisC gene product of Salmonella typhimurium.
KW   hisG; hisC; Salmonella typhimurium; mutant; auxotroph; detection;
KW   identification; mutagen; transition; transversion.
OS   Salmonella typhimurium.
PN   WO9510629-A.
PD   20-APR-1995.
PF   14-OCT-1994; 011617.
PR   15-OCT-1993; US-137627.
PA   (REGC ) UNIV CALIFORNIA.
PI   Ames BN, Gee P, Maron DM;
DR   WPI; 95-161813/21.
DR   N-PSDB; Q89706.
PT   Salmonella auxotrophic mutants for detecting mutagenic agents -
PT   comprise point mutation(s) in histidine genes hisG and hisC ; for
PT   identifying specific base changes induced by a mutagen
PS   Disclosure; Page 96-97; 135pp; English.
CC   Salmonella bacteria having substitutions in DNA sequences which
CC   render the bacteria non-selectable for a known Salmonella
CC   characteristic may be used for characterising specific DNA base
CC   substitutions mutations induced by a mutagen causing a reversion
CC   back to the wild type phenotype.  The mutant Salmonella are
CC   auxotrophic and only revert back to the wild type phenotype if the
CC   reverse mutation occurs at a specific site.  Base transitions/
CC   transversions are identified by exposing 2-4 derivative strains to a
CC   mutagen.  The strains are then cultured in medium selecting for a
CC   specific characteristic and the base transition/transversions are
CC   identified. His mutants were sequenced to identity the type and
CC   physical location of the mutations responsible for the his-
CC   phenotype.
SQ   Sequence   459 AA;

  Query Match           20.9%;  Score 70;  DB 1;  Length 459;
  Best Local Similarity  22.9%;  Pred. No. 4.86e+01;
  Matches    8;  Conservative   14;  Mismatches  12;  Indels   1;  Gaps   1;

Db    289 TACVEHVFDSETNYIILARFTASSSVFKSLWDQCII 323
          ::|   :|:  : |  ::|  ::|||: :| | :::
Qy      9 SICCSIIFNEASIYCQGKFINGTSVIPPL-DNSVL 42


RESULT   10
ID   W46733 standard; Protein; 610 AA.
AC   W46733;
DT   02-JUN-1998  (first entry)
DE   Amino acid sequence of endothelial leukocyte adhesion molecule-1.
KW   Endothelial leukocyte adhesion molecule-1; ELAM 1; inducible,
KW   cytokine, overexpression, cellular adhesion molecule;
KW   intracellular adhesion molecule-1; ICAM-1; antisense molecule;
KW   lipid mixture; Alzheimer's disease, multiple sclerosis;
KW   viral hepatitis; cholangitis; cardiac allograft rejection.
OS   Homo sapiens.
PN   WO9746671-A1.
PD   11-DEC-1997.
PF   22-MAY-1997; CA0347.
PR   30-MAY-1996; US-656758.
PA   (UYBR-) UNIV BRITISH COLUMBIA.
PI   Hope MJ, Kilmuk SR, Scherrer P, Semple SC;
DR   WPI; 98-042180/04.
DR   N-PSDB; V16208.
PT   Composition for treatment of conditions associated with
PT   overexpression of ICAM-1  - used to treat e.g. Alzheimer's disease,
PT   glomerulonephritis, rheumatoid arthritis etc.
PS   Disclosure; Pages 44-46; 81pp; English.
CC   The present sequence represents human endothelial leukocyte adhesion
CC   molecule-1 (ELAM-1). ELAM-1 is a membrane glycoprotein which is induced
CC   by a number of cytokines. Pathological conditions associated with the
CC   overexpression of cellular adhesion molecules, such as ELAM-1 or
CC   intracellular adhesion molecule-1 (ICAM-1), can be treated with the
CC   novel pharmaceutical composition of the invention. The composition
CC   comprises an effective amount of an ICAM-1 antisense molecule
CC   encapsulated in a lipid mixture, the lipid mixture comprising at
CC   least two members selected from phospholipids, sterols and
CC   cationic lipids. The composition is used in a method to treat
CC   pathological conditions associated with overexpression of ICAM-1, such
CC   as Alzheimer's disease, multiple sclerosis, viral hepatitis,
CC   cholangitis, cardiac allograft rejection, etc.
SQ   Sequence   610 AA;

  Query Match           20.9%;  Score 70;  DB 1;  Length 610;
  Best Local Similarity  28.6%;  Pred. No. 4.86e+01;
  Matches    6;  Conservative    9;  Mismatches   6;  Indels   0;  Gaps   0;

Db     40 AMTYDEASAYCQQRYTHLVAI 50
          :: ::||| ||| :: :  ::
Qy     13 SIIFNEASIYCQGKFINGTSV 33


RESULT   11
ID   R05538 standard; protein; 610 AA.
AC   R05538;
DT   24-OCT-1990  (first entry)
DE   Endothelial-leukocyte adhesion molecule (ELAM-1).
KW   Endothelial leukocyte adhesion molecules; ELAM 1; cytotoxin;
KW   cytokine; cancer; leukemia; ds.
OS   Homo sapiens.
PN   WO9005549-A.
PD   31-MAY-1990.
PF   14-NOV-1989; 005075.
PR   14-NOV-1988; US-270860.
PA   (BRIG ) Brigham & Womens Ho.
PI   Bevilacqua MP, Grimbrone MA;
DR   WPI; 90-193270/25.
DR   N-PSDB; Q04776.
PT   Antibodies against ELAM-1 to treat inflamation, infection,
PT   leukaemia -
PT   and for detecting expression of this endothelial leucocyte
PT   adhesion protein.
PS   Disclosure; 44pp; English.
CC   ELAM specific antibodies, attached to a suitable drug eg. a
CC   cytotoxin can direct the action of the drug to sites of
CC   inflammation, tissue damage and infection. The same Abs can be
CC   used to treat leukaemia, lymphoma, vasculitis and metastic spread
CC   of cancer cells. Labelled ELAM specific Abs can be used to detect
CC   expression of the adhesion molecules.
SQ   Sequence   610 AA;

  Query Match           20.9%;  Score 70;  DB 1;  Length 610;
  Best Local Similarity  28.6%;  Pred. No. 4.86e+01;
  Matches    6;  Conservative    9;  Mismatches   6;  Indels   0;  Gaps   0;

Db     40 AMTYDEASAYCQQRYTHLVAI 50
          :: ::||| ||| :: :  ::
Qy     13 SIIFNEASIYCQGKFINGTSV 33


RESULT   12
ID   R05494 standard; protein; 610 AA.
AC   R05494;
DT   22-OCT-1990  (first entry)
DE   Endothelial leukocyte adhesion molecule-1 (ELAM-1).
KW   Endothelial leukocyte adhesion molecule; ELAM-1; inflamation;
KW   leukaemia; lymphoma; metastasis; cancer; ds.
OS   Homo sapiens.
PN   WO9005786-A.
PD   31-MAY-1990.
PF   14-NOV-1989; 005073.
PR   14-NOV-1988; US-270873.
PA   (GEHO ) Gen Hospital Corp
```

PA (BRIG-) Brigham Womens Hosp.
PI Bevilacqua MP, Grimbrone MA, Seed B, Stengelin S;
DR WPI; 90-193436/25.
DR N-PSDB; Q04910.
PT Cloned genes expressing endothelial-leukocyte adhesion molecule-1 -
PT useful in treating inflamation, post reperfusion injury,
PT bacterial infection, vasculitis, leukaemia etc.
PS Claim 13; 36pp; English.
CC ELAM-1 is usetul in treatment of inflamation, post reperfusion
CC injury, vasculitis etc. It may also be useful in treatment and
CC diagnosis of leukaemia, lymphomas and metastic spread of tumour
CC cells, as well as in mediation of cell-cell interactions in
CC embryogenesis and organ development, in neoplasia, and wound and
CC tissue regeneration.
SQ Sequence 610 AA;

Query Match 20.9%; Score 70; DB 1; Length 610;
Best Local Similarity 28.6%; Pred. No. 4.86e+01;
Matches 6; Conservative 9; Mismatches 6; Indels 0; Gaps 0;

Db 30 AMTYDEASAYCQQRYTHLVAI 50
:: ::||| ||| :: : ::
Qy 13 SIIFNEASIYCQGKFINGTSV 33

RESULT 13
ID R08116 standard; protein; 610 AA.
AC R08116;
DT 27-FEB-1991 (first entry)
DE Endothelial cell-leucocyte adhesion molecule 1 from pCDM8 clone 6,
DE pSQ148 and pSQ149.
KW Endothelial cell-leucocyte adhesion molecule 1; ELAM1; antibodies;
KW molecule involved in leucocyte adhesion; MILA; inflamma ion.
OS Homo sapiens.
PN WO9013300-A.
PD 15-NOV-1990.
PF 27-APR-1990; U02457.
PR 28-APR-1989; US-345151.
PR 01-JUN-1989; US-359516.
PR 18-DEC-1989; US-452675.
PA (BIOG-) BIOGEN INC.
PI Hession C, Lobb RR, Goelz SE, Born L, Benjamin CD;
PI Rosa MD;
DR WPI; 90-361248/48.
DR N-PSDB; Q06686.
PT Endothelial cell adhesion mols. - MILAs and DNA encoding them and
PT inhibition-detection of binding of leukocytes to endothelial
PT cells
PS Disclosure; Fig 1 (A-E); 136pp; English.
CC The protein possesses a hydrophobic N-terminal sequence
CC characteristic of a signal sequence. It is predicted that the mature
CC N-terminal amino acid will be tryptophan at positi-n 22. The
CC extracellular domain of the polypeptide is ca. 554 amino acids including
CC the signal sequence and is followed by a hydrophobic transmembrane
CC region of 24 amino acids. The protein possesses a short, charged
CC cytoplasmic tail of 32 amino acids. It is noted that the protein
CC is cysteine-rich and contains eleven N-glycosylation sites.
CC ELAM can be used to identify molecules which inhibit binding of
CC leucocytes to cells, and such binding can be inhibited
CC (i.e. inflammation can be treated) using an ELAM (or fragment),
CC antibodies which recognise MILA, ELAM ligands or their fragments,
CC carbohydrates which bind to ELAM and antibodies which recognise ELAM.
CC See also Q06686-91.
SQ Sequence 610 AA;

Query Match 20.9%; Score 70; DB 1; Length 610;
Best Local Similarity 28.6%; Pred. No. 4.86e+01;
Matches 6; Conservative 9; Mismatches 6; Indels 0; Gaps 0;

Db 30 AMTYDEASAYCQQRYTHLVAI 50
:: ::||| ||| :: : ::
Qy 13 SIIFNEASIYCQGKFINGTSV 33

RESULT 14
ID W18839 standard; Protein; [illegible] AA.
AC W18839;
DT 05-JAN-1998 (first entry)
DE E-selectin.
KW E selectin; endothelial leukocyte adhesion molecule;
KW murine monoclonal antibody [illegible]; inflammation;
KW activated endothelium; treat; inflammation; microb
KW post-reperfusion injury; malignant tumour; vascul
KW vascular smooth-muscle cell proliferative disorde
KW chronic allograft rejection.
OS Homo sapiens.
PN US5632991-A.
PD 27-MAY-1997.
PF 14-NOV-1988; 270860.
PR 29-DEC-1994; US-365470.
PR 14-NOV-1988; US-270860.
PR 13-MAR-1992; US-850802.
PR 05-AUG-1993; US-102511.
PA (BGHM) BRIGHAM & WOMENS H SPITAL.
PI Gimbrone MA;
DR WPI; 97-297221/27.
DR N-PSDB; I70493.
PT Immuno:conjugates based on monoclonal antibody sp
PT E-selectin - for targetting therapeutic agent to
PT endothelium
PS Example 2; Column 35-38; 56pp; English.
CC This sequence is E selectin (formerly known as en
CC leukocyte adhesion molecule 1; ELAM-1). E-selectin
CC specific binding of murine monoclonal antibodies
CC developed against stimulated endothelial cells. M
CC of the invention comprise monoclonal antibody BBA
CC which binds to E selectin conjugated to a therape
CC immunoconjugate is used to selectively target the
CC to activated endothelium expressing E-selectin, e
CC inflammation, microbial infections, post-reperfus
CC tumours, vasculitis, vascular smooth-muscle cell
CC or acute or chronic allograft rejection.
SQ Sequence 610 AA;

Query Match 20.9%; Score 70; DB 1; Leng
Best Local Similarity 28.6%; Pred. No. 4.86e+01;
Matches 6; Conservative 9; Mismatches 6;

Db 30 AMTYDEASAYCQQRYTHLVAI 50
:: ::||| ||| :: : ::
Qy 13 SIIFNEASIYCQGKFINGTSV 33

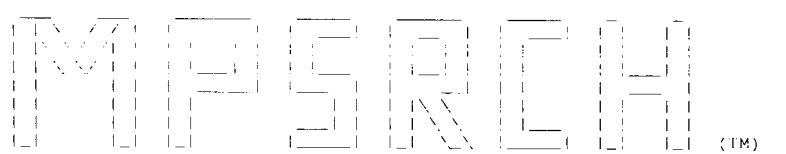
RESULT 15
ID R89280 standard; Protein; 4[illegible] AA.
AC R89280;
DT 08-APR-1996 (first entry)
DE CM-cellulase.
KW Tomato; CM-cellulase; endo-1,4 beta-D-glucanase;
KW ripening; probe; hybridization; cDNA library; Esc
KW fruit; cloning; antisense RNA; DNA cassette; cont
KW transgenic plant; fruit softening; crop impro emen
OS Lycopersicon esculentum.
PN WO9601555-A1.
PD 25 JAN 1996.
PF 07-JUL-1995; U08602.
PR 07-JUL 1994; US 271808.
PA (REGC) UNIV CALIFORNIA.
PI Bennett AB, Fischer RL, Giovannoni J, Lashbroo
DR WPI; 96-097388/10.
DR N-PSDB; Q99276.
PT Reducing fruit softening via antisense inhibitio
PT - by introducing into plant expression cassette
PT promoter and nucleic acid sequence of at least 2
PT endo-1,4-beta-glucanase gene
PS Disclosure; Page 42-44; 59pp; English.

CC The sequence represents a tomato ripening-associated CM-cellulase
CC (endo-1,4-beta-D-glucanase, EC-3.2.1.4), encoded by the tcl3 gene.
CC The encoding sequence has been isolated from a ripe tomato fruit
CC cDNA library in Escherichia coli using probes based on highly
CC conserved regions from bean and avocado glucanases. The cellulase
CC sequence may be used in construction of antisense DNA expression
CC cassettes, which may be inserted into vector constructs for
CC Agrobacterium-mediated gene transfer to e.g. tomato plants. The
CC antisense construct controls the activity of the gene during fruit
CC ripening, and reduces fruit softening by inhibiting glucanase
CC activity.
SQ Sequence 452 AA;

Query Match 20.6%; Score 69; DB 1; Length 452;
Best Local Similarity 36.4%; Pred. No. 5.95e+01;
Matches 8; Conservative 7; Mismatches 6; Indels 1; Gaps 1;

```
Db    115 SIVFKDNKAYSQ-KLVHGARTL 135
          ||:|::   | | |:::|:  |
Qy     13 SIIFNEASIYCQGKFINGTSVL 34
```

Search completed: Fri Aug 25 17:44:08 2000
Job time : 23 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 25 17:47:13 2000; MasPar time 2.47 Seconds
279.189 Million cell updates/sec

Tabular output not generated.

Title: >US-09-257-585-2
Description: (1-45) from US09257585.pep
Perfect Score: 335
Sequence: 1 MINKYPVESGCCSIFNEASYCQQCKFINCISVLPPLDNSVIGKR 45

Scoring table: PAM 150
Gap 11

Searched: 152433 seqs, 15329240 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:6_COMB 4:PCT_COMB 5:backfiles1

Statistics: Mean 22.710; Variance 85.811; scale 0.265

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	73	21.8	484	3	US-09-276-	Sequence 9, Applicatio	1.86e+01
2	73	21.8	484	2	US-08-252	Sequence 9, Applicatio	1.86e+01
3	72	21.5	463	2	US-08-162-	Sequence 9, Applicatio	2.27e+01
4	72	21.5	730	2	US-08-696-	Sequence 2, Applicatio	2.27e+01
5	71	21.2	540	1	US-08-307	Sequence 23, Applicati	2.72e+01
6	70	20.9	119	1	US-08-340-	Sequence 15, Applicati	3.37e+01
7	70	20.9	120	1	US-08-274-	Sequence 36, Applicati	3.37e+01
8	70	20.9	359	2	US-08-865-	Sequence 4, Applicatio	3.37e+01
9	70	20.9	359	1	US-08-137-	Sequence 4, Applicatio	3.37e+01
10	70	20.9	610	5	5217870-2	atent No. 5217870	3.37e+01
11	70	20.9	610	1	US-08-365-	Sequence 3, Applicatio	3.37e+01
12	69	20.6	452	1	US-08-271	Sequence 6, Applicatio	4.11e+01
13	69	20.6	452	1	US-08-484-	Sequence 6, Applicatio	4.11e+01
14	66	19.7	120	1	US-08-274	Sequence 37, Applicati	7.38e+01
15	65	19.4	128	2	US-08-928	Sequence 3, Applicatio	8.95e+01
16	65	19.4	490	2	US-08-724-	Sequence 6, Applicatio	8.95e+01
17	64	19.1	906	3	US-08-630-	Sequence 48, Applicati	1.08e+02
18	63	18.8	183	1	US-07-828-	Sequence 5, Applicatio	1.31e+02
19	62	18.5	54	2	US-08-693-	Sequence 42, Applicati	1.59e+02
20	62	18.5	394	1	US-07-603-	Sequence 11, Applicati	1.59e+02
21	62	18.5	417	1	US-08-277	Sequence 8, Applicatio	1.59e+02
22	62	18.5	417	1	US-08-430	Sequence 8, Applicatio	1.59e+02
23	62	18.5	417	1	US-07-649-	Sequence 8, Applicatio	1.59e+02
24	62	18.5	439	1	US-07-762-	Sequence 1,	[illegible]
25	62	18.5	500	1	US-08-307-	Sequence 2,	[illegible]
26	62	18.5	2516	3	US-08-474	Sequence 2,	[illegible]
27	61	18.2	120	2	US-08-928-	Sequence 1	[illegible]
28	61	18.2	378	2	US-08-333	Sequence 4,	[illegible]
29	61	18.2	378	2	US-08-333-	Sequence 4,	[illegible]
30	61	18.2	387	2	US-08-989-	Sequence 34	[illegible]
31	61	18.2	595	5	5234211-1	atent No. 5	[illegible]
32	61	18.2	619	2	US-08-810	Sequence 19	[illegible]
33	61	18.2	938	1	US-08-463	Sequence 7,	[illegible]
34	61	18.2	1129	5	5164380-6	atent No. 5	[illegible]
35	61	18.2	1174	1	US-07-828-	Sequence 1,	[illegible]
36	61	18.2	1174	1	US-08-157-	Sequence 1	[illegible]
37	61	18.2	1174	1	US-08-446-	Sequence 1,	[illegible]
38	61	18.2	1174	1	US-08-356	Sequence 2,	[illegible]
39	61	18.2	1174	4	PCT-US93-1	Sequence 1	[illegible]
40	61	18.2	1174	4	PCT-US93-1	Sequence 1	[illegible]
41	61	18.2	1174	4	PCT-US92-1	Sequence 1,	[illegible]
42	61	18.2	1177	5	5180668-2	atent No. 5	[illegible]
43	61	18.2	1177	1	US-08-040-	Sequence 1,	[illegible]
44	61	18.2	2165	1	US-08-514-	Sequence 2,	[illegible]
45	61	18.2	2165	4	PCT-US95-1	Sequence 2	[illegible]

ALIGNMENTS

RESULT 1
ID US-09-276-197-9 STANDARD; PRT; 484 AA.
XX
AC AAAAAA
XX
DT
XX
DE Sequence 9, Application US/09276197
XX
CC Sequence 9, Application US/09276197
CC Patent No. 6040428
CC GENERAL INFORMATION:
CC APPLICANT: Rollins, Scott
CC APPLICANT: Rother, Russell P.
CC APPLICANT: Evans, Mark J.
CC APPLICANT: Matis, Louis A.
CC TITLE OF INVENTION: P-SELECTIN E-SELECTIN
CC NUMBER OF SEQUENCES: 9
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Seth A. Fidel
CC STREET: 25 Science Park, Box 15
CC CITY: New Haven
CC STATE: Connecticut
CC COUNTRY: USA
CC ZIP: 06511
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: 3.5 inch, 720 kb storage
CC COMPUTER: PC compatible
CC OPERATING SYSTEM: DOS 6.
CC SOFTWARE: WordPerfect 6.
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/09/276,197
CC FILING DATE:
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/08/252,493
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Fidel, Seth A.
CC REGISTRATION NUMBER: 38,449
CC REFERENCE/DOCKET NUMBER: ALX-138
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (203) 772-1795
CC TELEFAX: (203) 772-1655
CC INFORMATION FOR SEQ ID NO: 9:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 484 amino acids

```
CC         TYPE:  amino acid
CC         STRANDEDNESS:  Single
CC         TOPOLOGY:  Linear
CC       MOLECULE TYPE:  cDNA to mRNA
CC         DESCRIPTION:  predicted amino acid sequence of
CC         DESCRIPTION:  Porcine E-selectin
SQ   SEQUENCE   484 AA; 52567 MW; 1238841 CN;

  Query Match            21.8%;  Score 73;  DB 3;  Length 484;
  Best Local Similarity  28.6%;  Pred. No. 1.86e+01;
  Matches    6;  Conservative    9;  Mismatches   6;  Indels   0;  Gaps   0;

Db      31 TMTFDDASAYCQQRYTHLVAI 51
           :: |::|| ||| :: :  ::
Qy      13 SIIFNEASIYCQGKFINGTSV 33


RESULT    2
ID   US-08-252-493C-9      STANDARD;      PRT;   484 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 9, Application US/08252493C
XX
CC    Sequence 9, Application US/08252493C
CC    Patent No. 5891645
CC     GENERAL INFORMATION:
CC       APPLICANT:  Rollins, Scott
CC       APPLICANT:  Rother, Russell P.
CC       APPLICANT:  Evans, Mark J.
CC       APPLICANT:  Matis, Louis A.
CC       TITLE OF INVENTION:  PORCINE E-SELECTIN
CC       NUMBER OF SEQUENCES:  9
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Seth A. Fidel
CC         STREET:  25 Science Park, Box 15
CC         CITY:  New Haven
CC         STATE:  Connecticut
CC         COUNTRY:  USA
CC         ZIP:  06511
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  3.5 inch, 750 Kb storage
CC         COMPUTER:  PC compatible
CC         OPERATING SYSTEM:  DOS 6.2
CC         SOFTWARE:  WordPerfect 6.0
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/252,493
CC         FILING DATE:  June 1, 1994
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:
CC         FILING DATE:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Fidel, Seth A.
CC         REGISTRATION NUMBER:  38,449
CC         REFERENCE/DOCKET NUMBER:  ALX-138
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (203) 776-1790
CC         TELEFAX:  (203) 772-4655
CC     INFORMATION FOR SEQ ID NO:  9:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  484 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  Single
CC         TOPOLOGY:  Linear
CC       MOLECULE TYPE:  cDNA to mRNA
CC         DESCRIPTION:  predicted amino acid sequence of
CC         DESCRIPTION:  Porcine E-selectin
SQ   SEQUENCE   484 AA; 52567 MW; 1238841 CN;

  Query Match            21.8%;  Score 73;  DB 2;  Length 484;
  Best Local Similarity  28.6%;  Pred. No. 1.86e+01;
  Matches    6;  Conservative    9;  Mismatches   6;  Indels   0;  Gaps   0;

Db      31 TMTFDDASAYCQQRYTHLVAI 51
           :: |::|| ||| :: :  ::
Qy      13 SIIFNEASIYCQGKFINGTSV 33


RESULT    3
ID   US-08-162-402B-9      STANDARD;      PRT;   463 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 9, Application US/08162402B
XX
CC    Sequence 9, Application US/08162402B
CC    Patent No. 5972337
CC     GENERAL INFORMATION:
CC       APPLICANT:  CERIANI, ROBERTO L.
CC       APPLICANT:  PETERSON, JERRY A.
CC       APPLICANT:  LAROCCA, DAVID J.
CC       TITLE OF INVENTION:  46 KDALTON HUMAN MILK FAT
CC       TITLE OF INVENTION:  GLOBULE (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
CC       NUMBER OF SEQUENCES:  29
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Pretty, Schroeder & Poplawski
CC         STREET:  444 South Flower St., 19th Floor
CC         CITY:  Los Angeles
CC         STATE:  CA
CC         COUNTRY:  USA
CC         ZIP:  90071
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette
CC         COMPUTER:  IBM Compatible
CC         OPERATING SYSTEM:  DOS
CC         SOFTWARE:  FastSEQ for Windows Version 2.0
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/162,402B
CC         FILING DATE:  03-DEC-1993
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:
CC         FILING DATE:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Amzel, Viviana
CC         REGISTRATION NUMBER:  30,930
CC         REFERENCE/DOCKET NUMBER:  PDS 36215
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  213-622-7700
CC         TELEFAX:  213-489-4210
CC         TELEX:
CC     INFORMATION FOR SEQ ID NO:  9:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  463 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  unknown
CC         TOPOLOGY:  unknown
CC       MOLECULE TYPE:  peptide
SQ   SEQUENCE   463 AA; 51416 MW; 1096958 CN;

  Query Match            21.5%;  Score 72;  DB 2;  Length 463;
  Best Local Similarity  29.4%;  Pred. No. 2.27e+01;
  Matches   10;  Conservative   10;  Mismatches  13;  Indels   1;  Gaps   1;

Db      16 CASGLFAASGDFCDSSLCLNGGTCLTGQDNDIYC 49
           | | :|  :: :|:: : :|| : |  || : |
Qy      11 CCSIIFNEASIYCQGKF-INGTSVLPPLDNSVLC 43


RESULT    4
ID   US-08-696-944-2      STANDARD;      PRT;   730 AA.
```

```
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application US/08696944
XX
CC    Sequence 2, Application US/08696944
CC    Patent No. 5981831
CC     GENERAL INFORMATION:
CC       APPLICANT:  Sumant CHENGAPPA
CC       APPLICANT:  Susan A. HELLYER
CC       APPLICANT:  John S. REID
CC       APPLICANT:  Jacqueline DE SILVA
CC       TITLE OF INVENTION:  No. 5981831el Exo-(1-4)-Beta-D Galactanase
CC       NUMBER OF SEQUENCES:  20
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Pillsbury Madison & Sutro, L.L.P.
CC         STREET:  1100 New York Avenue, N.W.
CC         CITY:  Washington
CC         STATE:  D.C.
CC         COUNTRY:  U.S.A.
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  3.5 inch disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  MS Word
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/696,944
CC         FILING DATE:  23-AUG-1996
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/GB95/00372
CC         FILING DATE:  23-FEB-1995
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  GB 9403423.8
CC         FILING DATE:  23-FEB-1994
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  730 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   730 AA; 81702 MW; 2877290 CN;

  Query Match           21.5%;  Score 72;  DB 2,  Length 730;
  Best Local Similarity  38.5%;  Pred. No. 2.27e+01;
  Matches   15;  Conservative    6;  Mismatches  17;  Indels   1;  Gaps   1;

Db    373 YKTESACAAFLANYNTDYSTQVKFGNGQYDLPPWSISIL 411
          |   |:| ::: |  : |  | || ||   |||   |:|
Qy      5 YPVKSICCSIIFNEASIYC-QCKFINGISVLPPLDNSVL 42


RESULT    5
ID   US-08-307-499-29     STANDARD;     PRT;   530 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 29, Application US/08307499
XX
CC    Sequence 29, Application US/08307499
CC    Patent No. 5651972
CC     GENERAL INFORMATION:
CC       APPLICANT:  Moyer, Richard W.
CC       APPLICANT:  Viquela, Eladio
CC       APPLICANT:  Gibbs, E.P.J.
CC       TITLE OF INVENTION:  Use of Recombinant Swine Poxvirus as a
CC       TITLE OF INVENTION:  Live Vaccine Vector
CC       NUMBER OF SEQUENCES:  60
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  David R. Saliwanchik
CC         STREET:  2421 N.W. 41st Street, Suite A-1
CC         CITY:  Gainesville
CC         STATE:  Florida
CC         COUNTRY:  U.S.A.
CC         ZIP:  32606
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/307,499
CC         FILING DATE:
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/908,241
CC         FILING DATE:  1-JUL-1992
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/908,630
CC         FILING DATE:  29-JUN-1992
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/342,212
CC         FILING DATE:  21-APR-1992
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Saliwanchik, David R.
CC         REGISTRATION NUMBER:  31,794
CC         REFERENCE/DOCKET NUMBER:  UF35.1-iwo-C1
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  904-375-8100
CC         TELEFAX:  904-372-5800
CC     INFORMATION FOR SEQ ID NO:  29:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  530 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   530 AA; 62649 MW; 1527568 CN;

  Query Match           21.2%;  Score 71;  DB 1;  Length
  Best Local Similarity  42.1%;  Pred. No. 2.77e+01;
  Matches   13;  Conservative    7;  Mismatches  12;

Db    415 MKLKQFPTAKSGISSIVYNCTYKIGGRIDIPHI 448
          | | :|: ||   |:: : | |   |:  :
Qy      1 MTNK YFV-KSTGGSIIFNEASIYCQGKFINGTSV 33


RESULT    6
ID   US 08 340 549A 15     STANDARD;     PRT;   [illegible] AA
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 15, Application US/08340549A
XX
CC    Sequence 15, Application US/08340549A
CC    Patent No. 5808025
CC     GENERAL INFORMATION:
CC       APPLICANT:  Tedder, Thomas F.
CC       APPLICANT:  Kansas, Geoffrey S.
CC       TITLE OF INVENTION:  CHIMERIC SELECTINS AS S
CC       TITLE OF INVENTION:  BLOCKING AGENTS FOR COMP
CC       NUMBER OF SEQUENCES:  28
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  FISH & NEAVE
CC         STREET:  1251 Avenue of the Americas
CC         CITY:  New York
CC         STATE:  New York
CC         COUNTRY:  USA
```

CC ZIP: 10020
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/340,539A
CC FILING DATE: 16 NOV 1994
CC CLASSIFICATION: 514
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/008,459
CC FILING DATE: 25 JAN 1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Gunnison, Jane
CC REGISTRATION NUMBER: 38,479
CC REFERENCE/DOCKET NUMBER: 08 194 CGN
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 212 596 9000
CC TELEFAX: 212-596-9090
CC INFORMATION FOR SEQ ID NO: 15:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 119 amino acids
CC TYPE: amino acid
CC STRANDEDNESS:
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 119 AA; 14118 MW; 73163 CN;

Query Match 20.9%; Score 70; DB 1; Length 119;
Best Local Similarity 28.6%; Pred. No. 3.37e+01;
Matches 6; Conservative 9; Mismatches 6; Indels 0; Gaps 0;

Db 9 AMTYDEASAYCQQRYTHLVAI 29
:: ::||| ||| :: : ::
Qy 13 SIIFNEASIYCQGKFINGTSV 33

RESULT 7
ID US-08-274-661B-36 STANDARD; PRT; 120 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 36, Application US/08274661B
XX
CC Sequence 36, Application US/08274661B
CC Patent No. 5593882
CC GENERAL INFORMATION:
CC APPLICANT: Erbe, David V.
CC APPLICANT: Lasky, Laurence A.
CC APPLICANT: Presta, Leonard G.
CC TITLE OF INVENTION: Selectin Variants
CC NUMBER OF SEQUENCES: 49
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Genentech, Inc.
CC STREET: 460 Point San Bruno Blvd
CC CITY: South San Francisco
CC STATE: California
CC COUNTRY: USA
CC ZIP: 94080
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WinPatin (Genentech)
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/274,661B
CC FILING DATE: 13-Jul-1994
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 07/956701
CC FILING DATE: 15/01/1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Dreger, Ginger R.
CC REGISTRATION NUMBER: 33,055
CC REFERENCE/DOCKET NUMBER: 761P1C1
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 415/225-3216
CC TELEFAX: 415/952-9881
CC TELEX: 910/371-7168
CC INFORMATION FOR SEQ ID NO: 36:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 120 amino acids
CC TYPE: Amino Acid
CC TOPOLOGY: Linear
SQ SEQUENCE 120 AA; 14189 MW; 73283 CN;

Query Match 20.9%; Score 70; DB 1; Length 120;
Best Local Similarity 28.6%; Pred. No. 3.37e+01;
Matches 6; Conservative 9; Mismatches 6; Indels 0; Gaps 0;

Db 9 AMTYDEASAYCQQRYTHLVAI 29
:: ::||| ||| :: : ::
Qy 13 SIIFNEASIYCQGKFINGTSV 33

RESULT 8
ID US-08-865-348-4 STANDARD; PRT; 359 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 4, Application US/08865348
XX
CC Sequence 4, Application US/08865348
CC Patent No. 5869258
CC GENERAL INFORMATION:
CC APPLICANT: Gee, Pauline
CC APPLICANT: Maron, Dorothy M.
CC APPLICANT: Ames, Bruce N.
CC TITLE OF INVENTION: A DETECTION SYSTEM FOR MUTAGENS THAT
CC TITLE OF INVENTION: ALSO IDENTIFIES MUTAGENIC CHANGES
CC NUMBER OF SEQUENCES: 21
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert,
CC ADDRESSEE: Attn: Karen S. Smith
CC STREET: 4 Embarcadero Center, Suite 3400
CC CITY: San Francisco
CC STATE: California
CC COUNTRY: United States
CC ZIP: 94111-4187
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/865,348
CC FILING DATE: 28-MAY-1997
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/137,627
CC FILING DATE: 15-OCT-1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Smith, Karen S.
CC REGISTRATION NUMBER: 31,426
CC REFERENCE/DOCKET NUMBER: A-58312-1/KSS
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415) 781-1989
CC TELEFAX: (415) 398-3249
CC TELEX: 910 277299
CC INFORMATION FOR SEQ ID NO: 4:
CC SEQUENCE CHARACTERISTICS:

```
CC         LENGTH:  359 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   359 AA; 39715 MW; 635027 CN;

  Query Match            20.9%;  Score 70;  DB 2;  Length 359;
  Best Local Similarity  22.9%;  Pred. No. 3.37e+01;
  Matches    8;  Conservative   14;  Mismatches  12;  Indels   1;  Gaps   1;

Db    289 IACVEHVFDSEINYILARFTASSSVFKSLWDQGII 323
          ::|   :|:  : |  ::|  ::|||: :| | :::
Qy      9 STCCSIIFNEASIYCQGKFINGTSVLPPL-DNSVL 42


RESULT   9
ID   US-08-137-627-4     STANDARD;     PRT;   359 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 4, Application US/08137627
XX
CC    Sequence 4, Application US/08137627
CC    Patent No. 5681737
CC     GENERAL INFORMATION:
CC       APPLICANT:  Goo, Pauline
CC       APPLICANT:  Maron, Dorothy M.
CC       APPLICANT:  Ames, Bruce N.
CC       TITLE OF INVENTION:  A DETECTION SYSTEM FOR MUTAGENS THAT
CC       TITLE OF INVENTION:  IDENTIFIES MUTAGENIC CHANGES
CC       NUMBER OF SEQUENCES:  44
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert,
CC         ADDRESSEE:  Attn: Karen S. Smith
CC         STREET:  4 Embarcadero Center, Suite 3400
CC         CITY:  San Francisco
CC         STATE:  California
CC         COUNTRY:  United States
CC         ZIP:  94111-4187
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/137,627
CC         FILING DATE:  15-OCT-1993
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Smith, Karen S.
CC         REGISTRATION NUMBER:  31,426
CC         REFERENCE/DOCKET NUMBER:  A-59412/KSS
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (415) 781-1989
CC         TELEFAX:  (415) 398-3249
CC         TELEX:  910 277299
CC     INFORMATION FOR SEQ ID NO:  4:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  359 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   359 AA; 39715 MW; 635027 CN;

  Query Match            20.9%;  Score 70;  DB 1;  Length 359;
  Best Local Similarity  22.9%;  Pred. No. 3.37e+01;
  Matches    8;  Conservative   14;  Mismatches  12;  Indels   1;  Gaps   1;

Db    289 IACVEHVFDSEINYILARFTASSSVFKSLWDQGII 323
          ::|   :|:  : |  ::|  ::|||: :| | :::
Qy      9 STCCSIIFNEASIYCQGKFINGTSVLPPL-DNSVL 42


RESULT   10
ID   5217870-2     STANDARD;     PRT:   610 AA.
XX
AC   xxxxxx
XX
DT
XX
DE  Patent No. 5217870
XX
CC   Patent No. 5217870
CC       APPLICANT: HESSION, CATHERINE A.;LOBB, ROY R.
CC       TITLE OF INVENTION: MONOCLONAL ANTIBODIES AGAI
CC       NUMBER OF SEQUENCES: 4
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  07/845,151
CC         FILING DATE: 28 APR 1989
CC   SEQ ID NO:2:
CC         LENGTH: 610
SQ   SEQUENCE   610 AA; 66655 MW; 1951337 CN;

  Query Match            20.9%;  Score 70;  DB 5;  Leng
  Best Local Similarity  28.6%;  Pred. No. 3.37e+01;
  Matches    6;  Conservative    9;  Mismatches   7;

Db     30 AMIYIDEASAYCQQRYCHLVA 53
          :: ::||| ||| :: : :  ::
Qy     13 SIIFNEASIYCQGKFINGTSV 34


RESULT   11
ID   US-08-365-470-3     STANDARD;     PRT:   610 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 3, Application US/08365470
XX
CC    Sequence 3, Application US/08365470
CC    Patent No. 5632991
CC     GENERAL INFORMATION:
CC       APPLICANT:  Gimbrone, Jr., Michael A.
CC       TITLE OF INVENTION:  Antibodies Specific
CC       TITLE OF INVENTION:  Thereof
CC       NUMBER OF SEQUENCES:  3
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FO
CC         STREET:  1100 New York Ave., NW
CC         CITY:  Washington
CC         STATE:  DC
CC         COUNTRY:  USA
CC         ZIP:  20005
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Vers
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/365,470
CC         FILING DATE:  Herewith
CC         CLASSIFICATION:  424
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 08/102,510
CC         FILING DATE:  05 AUG 1993
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/850,802
CC         FILING DATE:  13 MAR 1992
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Markowicz, Karen E.
CC         REGISTRATION NUMBER:  46,451
CC         REFERENCE/DOCKET NUMBER:  0627.1350003
```

```
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  202-371-2600
CC         TELEFAX:  202-371-2540
CC     INFORMATION FOR SEQ ID NO:  4:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  610 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
SQ   SEQUENCE   610 AA; 66681 MW; 1056017 CN;

  Query Match            20.9%;  Score 70;  DB 1;  Length 610;
  Best Local Similarity  28.6%;  Pred. No. 3.37e+01;
  Matches    6;  Conservative    9;  Mismatches   6;  Indels   0;  Gaps   0;

Db      30 AMTYDEASAYCQQRYTHLVA! 50
           :: ::||| ||| :: :  ::
Qy      13 SIIFNEASIYCQGKFINGTSV 33


RESULT   12
ID   US-08-271-883-6      STANDARD;      PRT;   452 AA.
XX
AC   XXXXXX
XX
DT
XX
DE   Sequence 6, Application US/08271883
XX
CC    Sequence 6, Application US/08271883
CC    Patent No. 5585545
CC     GENERAL INFORMATION:
CC       APPLICANT:  BENNETT, Alan B.
CC       APPLICANT:  FISCHER, Robert L.
CC       APPLICANT:  LASHBROOK, Coralie
CC       APPLICANT:  GIOVANNONI, James
CC       TITLE OF INVENTION:  ENDO-1,4-BETA-GLUCANASE GENES AND THEIR
CC       TITLE OF INVENTION:  USES
CC       NUMBER OF SEQUENCES:  10
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Townsend and Townsend Khourie and Crew
CC         STREET:  Steuart Street Tower, One Market Plaza
CC         CITY:  San Francisco
CC         STATE:  California
CC         COUNTRY:  US
CC         ZIP:  94105-1493
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/271,883
CC         FILING DATE:  07-JUL-1994
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/687,466
CC         FILING DATE:  18-APR-1991
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/511,417
CC         FILING DATE:  20-APR-1990
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Bastian, Kevin L.
CC         REGISTRATION NUMBER:  34,774
CC         REFERENCE/DOCKET NUMBER:  2307E-304-2
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (415) 543-9600
CC         TELEFAX:  (415) 543-5043
CC     INFORMATION FOR SEQ ID NO:  6:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  452 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   452 AA; 49935 MW; 1086902 CN;

  Query Match            20.6%;  Score 69;  DB 1;  Length 452;
  Best Local Similarity  36.4%;  Pred. No. 4.11e+01;
  Matches    8;  Conservative    7;  Mismatches   6;  Indels   1;  Gaps   1;

Db     115 SIVFKDNKAYSQ-KLVHGARTL 135
           ||:|::   | | |:::|:  |
Qy      13 SIIFNEASIYCQGKFINGTSVL 34


RESULT   13
ID   US-08-434-702-6      STANDARD;      PRT;   452 AA.
XX
AC   XXXXXX
XX
DT
XX
DE   Sequence 6, Application US/08434702
XX
CC    Sequence 6, Application US/08434702
CC    Patent No. 5554743
CC     GENERAL INFORMATION:
CC       APPLICANT:  Bennett, Alan B.
CC       APPLICANT:  Fischer, Robert L.
CC       APPLICANT:  Lashbrook, Coralie
CC       APPLICANT:  Giovanni, James
CC       TITLE OF INVENTION:  Endo-1,4-beta-Glucanase Genes and Their
CC       TITLE OF INVENTION:  Use in Plants
CC       NUMBER OF SEQUENCES:  10
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Townsend and Townsend and Crew
CC         STREET:  One Market Plaza, Steuart Street Tower
CC         CITY:  San Francisco
CC         STATE:  California
CC         COUNTRY:  USA
CC         ZIP:  94105-1492
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/434,702
CC         FILING DATE:  04-MAY-1995
CC         CLASSIFICATION:  536
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 08/271,883
CC         FILING DATE:  07-JUL-1994
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/687,466
CC         FILING DATE:  18-APR-1991
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/511,417
CC         FILING DATE:  20-APR-1990
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Bastian, Kevin L.
CC         REGISTRATION NUMBER:  34,774
CC         REFERENCE/DOCKET NUMBER:  2307E-304-30US
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (415) 543-9600
CC         TELEFAX:  (415) 543-5043
CC     INFORMATION FOR SEQ ID NO:  6:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  452 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   452 AA; 49935 MW; 1086902 CN;

  Query Match            20.6%;  Score 69;  DB 1;  Length 452;
  Best Local Similarity  36.4%;  Pred. No. 4.11e+01;
  Matches    8;  Conservative    7;  Mismatches   6;  Indels   1;  Gaps   1;
```

```
Db     115 SIVFKDNKAYSQ-KLVIGGARTL 135
           ||:|::   | | |:::|:  |
Qy      13 SIIFNEASIYCQGKFINGISVL 34


RESULT   14
ID   US-08-274-661B-37     STANDARD;     PRT;   120 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 37, Application US/08274661B
XX
CC    Sequence 37, Application US/08274661B
CC    Patent No. 5593882
CC     GENERAL INFORMATION:
CC       APPLICANT:  Erbe, David V.
CC       APPLICANT:  Lasky, Laurence A.
CC       APPLICANT:  Presta, Leonard G.
CC       TITLE OF INVENTION:  Selectin Variants
CC       NUMBER OF SEQUENCES:  39
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Genentech, Inc.
CC         STREET:  460 Point San Bruno Blvd
CC         CITY:  South San Francisco
CC         STATE:  California
CC         COUNTRY:  USA
CC         ZIP:  94080
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  WinPatin (Genentech)
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/274,661B
CC         FILING DATE:  13-Jul-1994
CC         CLASSIFICATION:  424
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  07/956701
CC         FILING DATE:  10/01/1992
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Dreger, Ginger R.
CC         REGISTRATION NUMBER:  33,055
CC         REFERENCE/DOCKET NUMBER:  761P1C1
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  415/225-3216
CC         TELEFAX:  415/952-9881
CC         TELEX:  910/371-7168
CC     INFORMATION FOR SEQ ID NO:  37:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  120 amino acids
CC         TYPE:  Amino Acid
CC         TOPOLOGY:  Linear
SQ   SEQUENCE   120 AA; 14135 MW; 72718 CN;

  Query Match           19.7%;  Score 66;  DB 1;  Length 120;
  Best Local Similarity  28.6%;  Pred. No. 7.3e-01;
  Matches    6;  Conservative    9;  Mismatches   6;  Indels   0;  Gaps   0;

Db       9 NMTYDEASAYCQQNYTHLVAI 29
           :: ::||| ||| :: :  ::
Qy      13 SIIFNEASIYCQGKFINGISV 33


RESULT   15
ID   US-08-928-926A-3     STANDARD;     PRT;   128 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 3, Application US/08928926A
XX
CC    Sequence 3, Application US/08928926A
CC    Patent No. 5925543
CC     GENERAL INFORMATION:
CC       APPLICANT:  Lal, Preeti
CC       APPLICANT:  Tang, Y. Tom
CC       APPLICANT:  Corley, Neil C.
CC       TITLE OF INVENTION:  NAD(H) DEHYDROGENASE BCP S
CC       NUMBER OF SEQUENCES:  4
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Incyte Pharmaceuticals, Inc.
CC         STREET:  3174 Porter Drive
CC         CITY:  Palo Alto
CC         STATE:  CA
CC         COUNTRY:  USA
CC         ZIP:  94304
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette
CC         COMPUTER:  IBM Compatible
CC         OPERATING SYSTEM:  DOS
CC         SOFTWARE:  FastSEQ for Windows Version 2.0
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/928,926A
CC         FILING DATE:
CC         CLASSIFICATION:  424
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:
CC         FILING DATE:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Billings, Lucy J.
CC         REGISTRATION NUMBER:  36,749
CC         REFERENCE/DOCKET NUMBER:  PF-0384 US
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  650/845-5500
CC         TELEFAX:  650/845-4166
CC         TELEX:
CC     INFORMATION FOR SEQ ID NO:  3:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  128 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       IMMEDIATE SOURCE:
CC         LIBRARY:  GenBank
CC         CLONE:  242
SQ   SEQUENCE   128 AA; 14124 MW; 98363 CN;

  Query Match           19.4%;  Score 65;  DB 2;  Length
  Best Local Similarity  28.3%;  Pred. No. 8.9e-01;
  Matches    8;  Conservative    7;  Mismatches   3;

Db      87 KYHVIIKPYIIVEKKPEEIPKGSILELGEVIPPM 120
           || | :  :|: : : :   :::  |:||:
Qy       4 KYPVKSTCCSIIFNEASIYLKRTFNSISVLPPL 37


Search completed: Fri Aug 25 17:47:21 2000
Job time : 8 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 25 17:44:25 2000; MasPar time 5.34 Seconds
497.689 Million cell updates/sec

Tabular output not generated.

Title: >US-09-257-585-2
Description: (1-45) from US09257585.pep
Perfect Score: 335
Sequence: 1 MENPVVPSIPGPVNLAGLTSLSLPFINCISVLPFLLNSVLPAR 45

Scoring table: PAM 150
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir64
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 33.123; Variance 51.723; scale 0.640

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	82	24.5	172	2	I47106	high-sulphur wool mat	4.52e-02
2	76	22.7	546	2	S66716	glutamate--tRNA ligas	4.75e-01
3	76	22.7	1048	2	S27763	Ca2+-transporting ATP	4.75e-01
4	75	22.4	240	2	A64403	peptidylprolyl isomer	6.96e-01
5	75	22.4	2743	2	S15760	genome polyprotein	6.96e-01
6	74	22.1	788	2	A46136	myosin-heavy-chain ki	1.02e+00
7	74	22.1	864	1	VCLJG4	env polyprotein - sim	1.02e+00
8	73	21.8	493	2	S48381	RRD1 protein - yeast	1.48e+00
9	73	21.8	482	2	JC5092	E-selectin - pig	1.48e+00
10	73	21.8	882	2	H70199	translation initiatio	1.48e+00
11	72	21.5	170	2	F70243	conserved hypothetica	2.14e+00
12	72	21.5	429	2	A36186	I factor 1 (transposo	2.14e+00
13	72	21.5	458	2	D70410	cytosolic axial filam	2.14e+00
14	72	21.5	463	1	A36479	milk fat globule memb	2.14e+00
15	72	21.5	668	2	I10575	hypothetical protein	2.14e+00
16	71	21.2	484	2	I39597	phosphoprotein phosph	3.09e+00
17	71	21.2	1849	1	RRWPEM	genome polyprotein -	3.09e+00
18	70	20.9	412	2	S31086	phosphoprotein phosph	4.45e+00
19	70	20.9	412	2	S24264	phosphoprotein phosph	4.45e+00
20	70	20.9	416	2	S29317	phosphoprotein phosph	4.45e+00
21	70	20.9	417	2	I04594	phosphoprotein phosph	4.45e+00
22	70	20.9	459	1	XNEBHC	histidinol-phosphate	4.45e+00
23	70	20.9	484	2	S26214	1-aminocyclopropane-1	4.45e+00
24	70	20.9	534	2	A57386	DNA helic	4.45e+00
25	70	20.9	610	2	A42695	endothelia	4.45e+00
26	70	20.9	611	2	A45343	endothelia	4.45e+00
27	70	20.9	664	2	I39544	protein-ta	4.45e+00
28	69	20.6	144	2	S34125	14-3-3e cha	6.00e+00
29	69	20.6	352	2	S42695	phosphopr	6.00e+00
30	69	20.6	416	2	S20225	phosphopr	6.00e+00
31	69	20.6	418	2	S20882	phosphopr	6.00e+00
32	69	20.6	425	2	A42543	phosphopr	6.00e+00
33	69	20.6	427	2	A42550	phosphopr	6.00e+00
34	69	20.6	470	2	I39654	1-aminocy	6.00e+00
35	69	20.6	484	2	S47418	profilaggri	6.00e+00
36	69	20.6	607	1	JQ1545	hemagglutin	6.00e+00
37	69	20.6	607	1	HMNZP3	hemagglut	6.00e+00
38	69	20.6	612	2	I37652	cellulase	6.00e+00
39	69	20.6	1361	2	T71403	hypothetic	6.00e+00
40	68	20.3	154	2	S70977	btpl prote	7.79e+00
41	68	20.3	326	2	I09544	phosphopr	7.79e+00
42	68	20.3	427	2	T34915	gas protei	7.79e+00
43	68	20.3	436	2	B71053	probable	7.79e+00
44	68	20.3	1252	2	S36016	oocyst wa	7.79e+00
45	68	20.3	1441	2	S57169	suilysin p	7.79e+00

ALIGNMENTS

RESULT 1
ENTRY I47106 #type complete
TITLE high-sulphur wool matrix protein bzA
ORGANISM #formal_name Ovis orientalis aries; D...
#common name domestic sheep
DATE 21-Feb-1997 #sequence_revision 21-Feb-...
13-Aug-1999
ACCESSIONS I47106
REFERENCE I47105
#authors Rogers, G.E.; Blackford, J.G.; Bickerst...
#journal Anim. Genet. (1994) 25:407-415
#title Polymorphism in two genes for Bz high-...
wool.
#cross-references MUID:95209345
#accession I47106
##status preliminary; translated from GB/...
##molecule_type DNA
##residues 1-172 ##label PRO
##cross-references GB:L35886; NID:g499872; PIDN:...
CLASSIFICATION #superfamily keratin high-sulfur matr...
SUMMARY #length 172 #molecular-weight 17884

Query Match 24.5%; Score 82; DB 2; Len...
Best Local Similarity 44.4%; Pred. No. 4.52e-...
Matches 12; Conservative 8; Mismatches 14; ...

Db 39 SIQTSCCQPIFIQIS-CCQPICQISCCQPICIQIS 7...
:: ::|| || :: :| || :: : | ||:|
Qy 6 FVKSTCCSIIFNEAIIYC-CKCINSICVLPP LDNS 4...

RESULT 2
ENTRY S66716 #type complete
TITLE glutamate--tRNA ligase (EC 6.1.1.17) ...
(Saccharomyces cerevisiae)
ALTERNATE_NAMES mitochondrial glutamyl-tRNA synthetas...
protein YOR335w
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 12-Jul-1996 #sequence_revision 12-Jul-...
16-Jul-1999
ACCESSIONS S66716; S59729
REFERENCE S66703
#authors Habbig, B.; Battenhorst, U.; Hollenb...
M.
#submission submitted to the Protein Sequence Da...
#accession S66716
##molecule_type DNA

```
      ##residues       1-536 ##label HAB
      ##cross-references EMBL:Z74775; NID:g1419824; PIDN:CAA99033.1;
                        PID:e251850; PID:g1419825; MIPS:YOL033w
      ##experimental_source strain S288C
REFERENCE         S56044
   #authors       Tzagoloff, A.A.; Shtanko, A.
   #submission    submitted to the EMBL Data Library, January 1995
   #accession     S59729
      ##molecule_type DNA
      ##residues       1-463,'L',465-536 ##label TZA
      ##cross-references EMBL:L39015; NID:g625181; PIDN:AAA61403.1;
                        PID:g625182
      ##experimental_source strain S273-10B
GENETICS
   #gene          SGD:MSE1
      ##cross-references SGD:S0005393; MIPS:YOL033w
   #map_position 15L
   #genome        nuclear
CLASSIFICATION    #superfamily glutamate--tRNA ligase; glutamine--tRNA ligase
                    homology
KEYWORDS          aminoacyl-tRNA synthetase; ATP; ligase; mitochondrion;
                    protein biosynthesis
FEATURE
   45-317             #domain glutamine -tRNA ligase homology #label EGL
SUMMARY           #length 536  #molecular-weight 61603  #checksum 8225

  Query Match             22.7%;  Score 76;  DB 2;  Length 536;
  Best Local Similarity   44.0%;  Pred. No. 4.75e-01;
  Matches    11;  Conservative    8;  Mismatches   4;  Indels   2;  Gaps   2;

Db      11 YCSPSKLIKGVGISPIKKSILSKK 34
           ||  :|:|:| : |:|| :|:| |:
Qy      22 YC QGKFINGTSVLPPLDNSVLCKK 45


RESULT    3
ENTRY             S27763      #type complete
TITLE             Ca2+-transporting ATPase (EC 3.6.1.38) LCA1 - tomato
ORGANISM          #formal_name Lycopersicon esculentum #common_name tomato
DATE              17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change
                    22-Jun-1999
ACCESSIONS        A46284; S27763
REFERENCE         A46284
   #authors       Wimmers, L.E.; Ewing, N.N.; Bennett, A.B.
   #journal       Proc. Natl. Acad. Sci. U.S.A. (1992) 89:9205-9209
   #title         Higher plant Ca(2+)-ATPase: primary structure and regulation
                    of mRNA abundance by salt.
   #cross-references MUID:93028433
   #accession     A46284
      ##status         preliminary; not compared with conceptual translation
      ##molecule_type nucleic acid
      ##residues       1-1048 ##label WI2
      ##cross-references EMBL:M96324; NID:g170377; PIDN:AAA34138.1;
                        PID:g170378
GENETICS
   #introns       490:2  [illegible]
CLASSIFICATION    #superfamily Na+/K+ transporting ATPase alpha chain; ATPase
                    nucleotide-binding domain homology
KEYWORDS          ATP; hydrolase
FEATURE
   618-794            #domain ATPase nucleotide-binding domain homology #label
                        ATN
SUMMARY           #length 1048  #molecular-weight 116127  #checksum 6884

  Query Match             22.7%;  Score 76;  DB 2;  Length 1048;
  Best Local Similarity   28.6%;  Pred. No. 4.75e-01;
  Matches    10;  Conservative   16;  Mismatches   7;  Indels   2;  Gaps   2;

Db     416 MDANELLMAEICAIC-NDACVPCDCRLFKATG-LP 448
           |  :  :::  |:|  |:|:::|:|::::::|: ||
Qy       1 MTNKYPVKSICCSIFNEASIYCQGKFINGTSVLP 35


RESULT    4
ENTRY             A64403      #type complete
TITLE             peptidylprolyl isomerase (EC 5.2.1.8)  Methanococcus
                    jannaschii
ALTERNATE_NAMES   peptidyl-prolyl cis trans isomerase, rotamase
ORGANISM          #formal_name Methanococcus jannaschii
DATE              13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change
                    13-Sep-1998
ACCESSIONS        A64403
REFERENCE         A64300
   #authors       Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann,
                    R.D.; Sutton, G.G.; Blake, J.A.; FitzGerald, L.M.; Clayton,
                    R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.;
                    Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.;
                    Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
                    Scott, J.L.; Geoghagen, N.S.M.; Weidman, J.F.; Fuhrmann,
                    J.L.; Nguyen, D.; Utterback, T.R.; Kelley, J.M.; Peterson,
                    J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts,
                    K.M.; Hurst, M.A.; Kaine, B.P.; Borodovsky, M.; Klenk,
                    H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
   #journal       Science (1996) 273:1058-1073
   #title         Complete genome sequence of the methanogenic archaeon,
                    Methanococcus jannaschii.
   #cross-references MUID:96337999
   #accession     A64403
      ##status         preliminary; nucleic acid sequence not shown;
                        translation not shown
      ##molecule_type DNA
      ##residues       1-240 ##label BUL
      ##cross-references GB:U67526; GB:L77117; NID:g1591508; PID:g1591512;
                        TIGR:MJ0825; PID:g1510883
GENETICS
   #map_position REV/47317-/46595
   #start_codon  TTG
CLASSIFICATION    #superfamily Methanococcus jannaschii peptidylprolyl
                    isomerase
KEYWORDS          cis trans isomerase; cyclosporin A binding
SUMMARY           #length 240  #molecular-weight 26976  #checksum 4054

  Query Match             22.4%;  Score 75;  DB 2;  Length 240;
  Best Local Similarity   41.4%;  Pred. No. 6.96e-01;
  Matches    12;  Conservative    5;  Mismatches  11;  Indels   1;  Gaps   1;

Db      43 IYNPAMIYGPVAIFAGEGQVLPGLDEAIL 71
           |:| | ||   : |     ||| ||:::|
Qy      15 IFNEASIYCQ-GKFINGTSVLPPLDNSVL 42


RESULT    5
ENTRY             S15760      #type complete
TITLE             genome polyprotein - murine hepatitis virus (strain A59)
ALTERNATE_NAMES   protein 1b
CONTAINS          RNA-directed RNA polymerase (EC 2.7.7.48)
ORGANISM          #formal_name murine hepatitis virus, MHV
DATE              30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
                    29-May-1998
ACCESSIONS        S15760, S08652
REFERENCE         S15760
   #authors       Bredenbeek, P.J.; Pachuk, C.J.; Noten, A.F.H.; Charite, J.;
                    Luytjes, W.; Weiss, S.R.; Spaan, W.J.M.
   #journal       Nucleic Acids Res. (1990) 18:1825-1832
   #title         The primary structure and expression of the second open
                    reading frame of the polymerase gene of the coronavirus
                    MHV-A59; a highly conserved polymerase is expressed by an
                    efficient ribosomal frameshifting mechanism.
   #cross-references MUID:90245573
   #accession     S15760
      ##molecule_type genomic RNA
      ##residues       1-2733 ##label BRE
      ##cross-references EMBL:X51939; NID:g58974; PID:e17193; PID:g1334829
CLASSIFICATION    #superfamily infectious bronchitis virus RNA-directed RNA
                    polymerase
KEYWORDS          nucleotidyltransferase; RNA biosynthesis
```

```
SUMMARY           #length 2733  #molecular-weight 309219  #checksum 5999

  Query Match              22.4%;  Score 75;  DB 2,  Length 2733;
  Best Local Similarity    26.2%;  Pred. No. 6.96e-01;
  Matches    11;  Conservative    14;  Mismatches   14;  Indels    3,  Gaps    3;

Db    908 DSYSVIISTCIDQQKFIDEIEY-KNMYLR-SAVLQSVGACVVC 947
          : ||:|  ||| :  | : :||   :::  ::|| ::   |:|
Qy      3 NKYPVK-STCCSIIFNEASIYCQGKFINCTSVLPPLDNSVLC 43


RESULT    6
ENTRY             A46136      #type complete
TITLE             myosin-heavy chain kinase (EC 2.7.1.123)  slime mold
                    (Dictyostelium discoideum)
ORGANISM          #formal_name Dictyostelium discoideum
DATE              21-Sep-1993 #sequence_revision 25 Apr 1997 #text_change
                    17-Apr-1998
ACCESSIONS        A46136
REFERENCE         A46136
   #authors       Ravid, S.; Spudich, J.A.
   #journal       Proc. Natl. Acad. Sci. U.S.A. (1992) 89:5877-5881
   #title         Membrane-bound Dictyostelium myosin heavy chain kinase: a
                    developmentally regulated substrate-specific member of the
                    protein kinase C family.
   #cross-references MUID:92335200
   #contents      AX3
   #accession     A46136
      ##status          preliminary
      ##molecule_type mRNA
      ##residues        1-783 ##label RAV
      ##cross-references GB:M93394; NID:g167827
      ##note            sequence extracted from NCBI backbone (NCBIN:108552,
                         NCBIP:108553); this ORF is not annotated in GenBank
                         entry DDIMHCK, release 103
CLASSIFICATION    #superfamily protein kinase C zinc-binding repeat homology
KEYWORDS          phosphotransferase
FEATURE
   88-138             #domain protein kinase C zinc-binding repeat homology
                        #label KZ2
SUMMARY           #length 783  #molecular-weight 89154  #checksum 4970

  Query Match              22.1%;  Score 74;  DB 2;  Length 783;
  Best Local Similarity    60.0%;  Pred. No. 1.02e+00;
  Matches    12;  Conservative    2;  Mismatches    5;  Indels    1;  Gaps    1;

Db    487 QTKEIVGVTVLPHLDD-VLC 505
          | | | | :||| ||: |||
Qy     24 QGKFINGTSVLPPLDNSVLC 43


RESULT    7
ENTRY             VCLJG4      #type complete
TITLE             env polyprotein - simian immunodeficiency virus (African
                    green monkey isolate)
ORGANISM          #formal_name simian immunodeficiency virus, SIV
DATE              30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change
                    24-Oct-1997
ACCESSIONS        G30045
REFERENCE         A30045
   #authors       Fukasawa, M.; Miura, T.; Hasegawa, A.; Morikawa, S.;
                    Tsujimoto, H.; Miki, K.; Kitamura, T.; Hayami, M.
   #journal       Nature (1988) 333:457-461
   #title         Sequence of simian immunodeficiency virus from African green
                    monkey, a new member of the HIV/SIV group.
   #cross-references MUID:88242906
   #accession     G30045
      ##molecule_type DNA
      ##residues        1-864 ##label FUK
      ##cross-references EMBL:X07805
GENETICS
   #gene          env
CLASSIFICATION    #superfamily type E retrovirus env polyprotein
KEYWORDS          AIDS; capsid protein; coat protein; gp
                    polyprotein; transmembrane protein
SUMMARY           #length 864  #molecular-weight 96414

  Query Match              21.1%;  Score 74;  DB 1;  Leng
  Best Local Similarity    44.4%;  Pred. No. 1.02e+
  Matches    11;  Conservative    4;  Mismatches   9;

Db    178 KENYYSVVWNQAEIY[illegible]
          |    |:: |:| ||| : ::|
Qy      8 KSTCCSIIFNEASIYCQGKFINGTS 32


RESULT    8
ENTRY             S48381      #type complete
TITLE             RRD1 protein - yeast (Saccharomyces c
ALTERNATE_NAMES   protein YIL154w
ORGANISM          #formal_name Saccharomyces cerevisiae
DATE              02-Dec-1994 #sequence_revision 02-Dec
                    13-Sep-1996
ACCESSIONS        S48381
REFERENCE         S48410
   #authors       Churcher, C.
   #submission    submitted to the EMBL Data Library, Se
   #accession     S48381
      ##molecule_type DNA
      ##residues       1-993 ##label CHU
      ##cross-references GB:Z47047; EMBL:Z38059; NID:g
                        MIPS:YIL154w
GENETICS
   #gene          RRD1
   #map_position 9L
SUMMARY           #length 993  #molecular-weight 45082

  Query Match              21.8%;  Score 73;  DB 2;  Leng
  Best Local Similarity    42.0%;  Pred. No. 1.48e+00;
  Matches     8;  Conservative    11;  Mismatches    4;

Db    242 YKDTNFYCQGINFINEVK-MGHFEE 265
          ::::::|||| :||:    : |:::
Qy     16 FNEASIYCQG-KFINGTSVLPPLDN 39


RESULT    9
ENTRY             JC5092      #type complete
TITLE             E-selectin - pig
ORGANISM          #formal_name Sus scrofa domestica #or
DATE              31-Jan-1997 #sequence_revision 31-Jan
                    16-Jul 1999
ACCESSIONS        JC5092
REFERENCE         JC5092
   #authors       Winkler, H.; Preston, C.; Csizmadia,
                    Anrather, J.; Bach, F.H.
   #journal       Gene (1996) 176:67-72
   #title         The intron-exon structure of the por
                    gene.
   #cross-references MUID:97079411
   #contents      endothelial cells
   #accession     JC5092
      ##molecule_type DNA
      ##residues       1-482 ##label WIN
      ##cross-references GB:U47521; NID:g1052974; EMBL
COMMENT     This protein is a member of the selectin fa
              molecules.
GENETICS
   #[illegible]
CLASSIFICATION    #superfamily complement factor H repe
                    lectin homology
FEATURE
   13-139             #domain C type lectin homology #
   181-235            #domain complement factor H repe
   240-298            #domain complement factor H repe
   303-361            #domain complement factor H repe
```

```
   366-420             #domain complement factor H repeat homology #label FH4
SUMMARY             #length 482  #molecular weight 52341  #checksum 7729

  Query Match            21.8%;  Score 73;  DB 2;  Length 482;
  Best Local Similarity  28.6%;  Pred. No. 1.48e+00;
  Matches     6;  Conservative     0;  Mismatches    6;  Indels    0;  Gaps    0;

Db      31 TMTFDDASAYCQQRYTHLVAI 51
           :: |:|| ||| :: :  ::
Qy      13 SIIFNEASIYCQGKFINGTSV 33


RESULT   10
ENTRY               H70199      #type complete
TITLE               translation initiation factor IF-2 - Lyme disease spirochete
ORGANISM            #formal_name Borrelia burgdorferi #common_name Lyme disease
                      spirochete
DATE                13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change
                      16-Jul-1999
ACCESSIONS          H70199
REFERENCE           A70100
   #authors         Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.;
                      Clayton, R.; Lathigra, R.; White, O.; Ketchum, K.A.;
                      Dodson, R.; Hickey, E.K.; Gwinn, M.; Dougherty, B.; Tomb,
                      J.F.; Fleischmann, R.D.; Richardson, D.; Peterson, J.;
                      Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.;
                      Vugt, R.V.; Palmer, N.; Adams, M.D.; Gocayne, J.; Weidman,
                      J.; Utterback, T.; Watthey, L.; McDonald, L.; Artiach, P.;
                      Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst,
                      K.; Roberts, K.; Hatch, B.; Smith, H.O.; Venter, J.C.
   #journal         Nature (1997) 390:580-586
   #title           Genomic sequence of a Lyme disease spirochete, Borrelia
                      burgdorferi.
   #cross-references MUID:98065943
   #accession       H70199
      ##status          preliminary; nucleic acid sequence not shown;
                         translation not shown
      ##molecule_type DNA
      ##residues       1-882 ##label KLE
      ##cross-references GB:AE000783; NID:g2690738; PID:g2690747;
                         TIGR:BB0801
      ##experimental_source strain B31
CLASSIFICATION      #superfamily translation elongation factor Tu homology
KEYWORDS            GTP binding; P-loop
FEATURE
   383-492             #domain translation elongation factor Tu homology #label
                         ETU\
   389-396             #region nucleotide-binding motif A (P-loop)\
   489-492             #region GTP-binding NKXD motif\
   525-527             #region GTP binding SAK/L motif\
   395,396,416,489,
   490,492,525         #binding_site Mg-GTP (Lys, Thr, Thr, Asn, Lys, Asp, Ser)
                         #status predicted
SUMMARY             #length 882  #molecular weight 97795  #checksum 854

  Query Match            21.8%;  Score 73;  DB 2;  Length 882;
  Best Local Similarity  25.0%;  Pred. No. 1.48e+00;
  Matches     6;  Conservative    11;  Mismatches    7;  Indels    0;  Gaps    0;

Db     569 RGIVCSVIIEDGTLYVGDSEVGGA 592
           ::  ||:|:::::|    |: |:
Qy       8 KSTCCSIIFNEASIYCQGKFINGT 31


RESULT   11
ENTRY               F70243      #type complete
TITLE               conserved hypothetical protein BB131 - Lyme disease
                      spirochete plasmid lp28-4
ORGANISM            #formal_name Borrelia burgdorferi #common_name Lyme disease
                      spirochete
DATE                13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change
                      05-Jun-1998
ACCESSIONS          F70243
REFERENCE           A70100
   #authors         Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.;
                      Clayton, R.; Lathigra, R.; White, O.; Ketchum, K.A.;
                      Dodson, R.; Hickey, E.K.; Gwinn, M.; Dougherty, B.; Tomb,
                      J.F.; Fleischmann, R.D.; Richardson, D.; Peterson, J.;
                      Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.;
                      Vugt, R.V.; Palmer, N.; Adams, M.D.; Gocayne, J.; Weidman,
                      J.; Utterback, T.; Watthey, L.; McDonald, L.; Artiach, P.;
                      Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst,
                      K.; Roberts, K.; Hatch, B.; Smith, H.O.; Venter, J.C.
   #journal         Nature (1997) 390:580-586
   #title           Genomic sequence of a Lyme disease spirochaete, Borrelia
                      burgdorferi.
   #cross-references MUID:98065943
   #accession       F70243
      ##status          preliminary; nucleic acid sequence not shown;
                         translation not shown
      ##molecule_type DNA
      ##residues       1-170 ##label KLE
      ##cross-references GB:AE000789; NID:g2690079; PID:g2690091; TIGR:BB131
      ##experimental_source strain B31
GENETICS
   #genome          plasmid
SUMMARY             #length 170  #molecular weight 18227  #checksum 8312

  Query Match            21.5%;  Score 72;  DB 2;  Length 170;
  Best Local Similarity  39.1%;  Pred. No. 2.14e+00;
  Matches     9;  Conservative     6;  Mismatches    7;  Indels    1;  Gaps    1;

Db      67 FVQGDYIGGGAVLGSQLLGGILC 89
           : || :| | :|| : |  ::||
Qy      22 YCQGKFINGTSVL-PPLINSVLC 43


RESULT   12
ENTRY               A36186      #type complete
TITLE               I factor 1 (transposon) - fruit fly (Drosophila teissieri)
ORGANISM            #formal_name Drosophila teissieri
DATE                14-Dec-1990 #sequence_revision 14-Dec-1990 #text_change
                      24-Sep-1998
ACCESSIONS          A36186
REFERENCE           A36186
   #authors         Abad, P.; Vaury, C.; Pelisson, A.; Chaboissier, M.C.;
                      Busseau, I.; Bucheton, A.
   #journal         Proc. Natl. Acad. Sci. U.S.A. (1989) 86:8887-8891
   #title           A long interspersed repetitive element--the I factor of
                      Drosophila teissieri--is able to transpose in different
                      Drosophila species.
   #cross-references MUID:90046894
   #accession       A36186
      ##status          preliminary
      ##molecule_type DNA
      ##residues       1 429 ##label ABA
      ##cross-references GB:M28878; NID:g157750; PID:g950415
GENETICS
   #gene            FlyBase:Dtei/I-element
      ##cross-references FlyBase:FBgn0013017
SUMMARY             #length 429  #molecular weight 48833  #checksum 1942

  Query Match            21.5%;  Score 72;  DB 2;  Length 429;
  Best Local Similarity  26.9%;  Pred. No. 2.14e+00;
  Matches     7;  Conservative     9;  Mismatches    9;  Indels    1;  Gaps    1;

Db      97 TLNFSKGVIYCNDLRYIDEDTILQEL 122
           :: |: : |||   ::|:  ::|  |
Qy      13 SIIFNEASIYCQG-KFINGTSVLPPL 37


RESULT   13
ENTRY               D70410      #type complete
TITLE               cytosolic axial filament protein - Aquifex aeolicus
ORGANISM            #formal_name Aquifex aeolicus
DATE                08 May 1998 #sequence_revision 08 May-1998 #text_change
```

```
                  02-Jul-1998
ACCESSIONS        D70410
REFERENCE         A70300
   #authors       Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.;
                    Lenox, A.L.; Graham, D.E.; Overbeek, R.; Snead, M.A.;
                    Keller, M.; Aujay, M.; Huber, R.; Feldman, R.A.; Short,
                    J.M.; Olson, G.J.; Swanson, R.V.
   #journal       Nature (1998) 392:353-358
   #title         The complete genome of the hyperthermophilic bacterium
                    Aquifex aeolicus.
   #cross-references MUID:98196666
   #accession     D70410
      ##status         preliminary; nucleic acid sequence not shown;
                         translation not shown
      ##molecule_type DNA
      ##residues       1-458 ##label AQF
      ##cross-references GB:AE000732; NID:g2983704; PID:g2983706; GB:AE000657
      ##experimental_source strain VF5
GENETICS
   #gene          cafA
SUMMARY           #length 458  #molecular-weight 52508  #checksum 3184

  Query Match           21.5%;  Score 72;  DB 2;  Length 458;
  Best Local Similarity  30.2%;  Pred. No. 2.14e+00;
  Matches   13;  Conservative   11;  Mismatches  15;  Indels   4;  Gaps   4;

Db     364 MTKKWESESIY-FVI-GETCPIPCQGKGFVKSKEFILHTLEKAL 404
           ||:|   |    :|  |:   |||| |:::   :|  |:::
Qy       1 MTNKYPVKSTCCSIIFNEASIYCQGK-FINGTS-VLPPLDNSV 41


RESULT   14
ENTRY             A36479      #type complete
TITLE             milk fat globule membrane protein - mouse
ORGANISM          #formal_name Mus musculus #common_name house mouse
DATE              10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change
                    10-Sep-1999
ACCESSIONS        A36479
REFERENCE         A36479
   #authors       Stubbs, J.D.; Lekutis, C.; Singer, K.L.; Bui, A.; Yuzuki, D.;
                    Srinivasan, U.; Parry, G.
   #journal       Proc. Natl. Acad. Sci. U.S.A. (1990) 87:8417-8421
   #title         cDNA cloning of a mouse mammary epithelial cell surface
                    protein reveals the existence of epidermal growth
                    factor-like domains linked to factor VIII-like sequences.
   #cross-references MUID:91046008
   #accession     A36479
      ##status         preliminary
      ##molecule_type mRNA
      ##residues       1-463 ##label SIU
      ##cross-references GB:M38337; NID:g199142; PIDN:AAA39534.1; PID:g199143
CLASSIFICATION    #superfamily milk fat globule protein; discoidin I
                    amino-terminal homology; EGF homology
KEYWORDS          membrane protein
FEATURE
   28-60               #domain EGF homology #label EG1\
   68-107              #domain EGF homology #label EG2\
   147-304             #domain discoidin I amino-terminal homology #label DN1\
   307-464             #domain discoidin I amino-terminal homology #label DN2
SUMMARY           #length 463  #molecular-weight 51465  #checksum 8337

  Query Match           21.5%;  Score 72;  DB 1;  Length 463;
  Best Local Similarity  29.4%;  Pred. No. 2.14e+00;
  Matches   10;  Conservative   10;  Mismatches  13;  Indels   1;  Gaps   1;

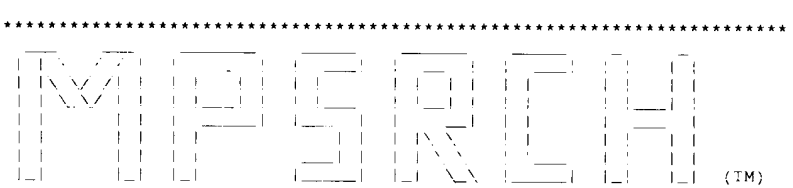
Db      16 CASGLFAASGDFCDSSLCLNGGTCLTGQDNDIYC 49
           | | :|  :: :|:: : :|| : |   || : |
Qy      11 CCSIIFNEASIYCQGKF-INGTSVLPPLDNSVLC 43


RESULT   15
ENTRY             T10575      #type complete
TITLE             hypothetical protein F25E4.170 - Arabidopsis thaliana
ORGANISM          #formal_name Arabidopsis thaliana #com
                    cross
DATE              16-Jul-1999 #sequence_revision 16-Jul
                    20-Sep-1999
ACCESSIONS        T10575
REFERENCE         Z16533
   #authors       Bevan, M.; Lennard, N.; Quail, M.; Har
                    M.A.; Barrell, B.G.; Hancroft, I.; M
                    K.F.X.; Lemcke, K.; Schueller, C.
   #submission    submitted to the Protein Sequence Data
   #accession     T10575
      ##status         preliminary
      ##molecule_type DNA
      ##residues       1-668 ##label BEV
      ##cross-references EMBL:AL0[illegible]; GSPDB:GN00[illegible]
      ##experimental_source cultivar Columbia; BAC clo
GENETICS
   #gene          ATSP:F25E4.170
   #map_position 4
SUMMARY           #length 668  #molecular weight 76017

  Query Match           21.5%;  Score 72;  DB 2;  Len
  Best Local Similarity  29.4%;  Pred. No. 2.14e+0
  Matches    8;  Conservative   10;  Mismatches  10;

Db      20 MPCNYLTRADCCGRFCHISGTYTLSTSFL 49
           |  :|  :: (|: : |  ::  ::  :|
Qy       1 MTNKYIVKETCCSII INDASIYCQG KFI 28


Search completed: Fri Aug 25 17:44:4, 2000
Job time : 17 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 25 17:45:00 2000; MasPar time 3.30 Seconds
421.995 Million cell updates/sec

Tabular output not generated.

Title: >US-09-257-585-2
Description: (1-45) from US09257585.pop
Perfect Score: 335
Sequence: 1 MINKYPVKSTCCSIIFNEASIYCQGKFINGISVLPPLDNSVLCKR 45

Scoring table: PAM 150
Gap 11

Searched: 85661 seqs, 30989116 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 33.849; Variance 47.020; scale 0.720

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	78	23.3	664	1	PDI3_MOUSE	PROTEIN ARGININE DEIMI	5.40e-02
2	76	22.7	536	1	SYEM_YEAST	GLUTAMYL-TRNA SYNTHETA	1.26e-01
3	75	22.4	231	1	FKB2_METJA	PUTATIVE FKBP-TYPE PEP	1.92e-01
4	75	22.4	611	1	LEM2_CANFA	E-SELECTIN PRECURSOR (	1.92e-01
5	75	22.4	616	1	SSH2_CAEEL	SEC-2 PROTEIN.	1.92e-01
6	75	22.4	2243	1	RRPB_CVMA5	RNA-DIRECTED RNA POLYM	1.92e-01
7	74	22.1	783	1	KMHC_DICDI	MYOSIN HEAVY CHAIN KIN	2.91e-01
8	74	22.1	865	1	ENV_SIVAT	ENVELOPE POLYPROTEIN G	2.91e-01
9	73	21.8	393	1	YIP3_YEAST	HYPOTHETICAL 45.1 KDA	4.40e-01
10	73	21.8	484	1	LEM2_PIG	E-SELECTIN PRECURSOR (	4.40e-01
11	73	21.8	882	1	IF2_BOBBU	TRANSLATION INITIATION	4.40e-01
12	72	21.5	463	1	MFGM_MOUSE	LACTADHERIN PRECURSOR	6.61e-01
13	72	21.5	570	1	SP26_CAEEL	SPERMATOCYTE PROTEIN S	6.61e-01
14	71	21.2	530	1	VC04_SPVKA	PROTEIN C4.	9.89e-01
15	71	21.2	1839	1	POLR_EPMV	RNA REPLICASE POLYPROT	9.89e-01
16	70	20.9	312	1	PP12_ARATH	SERINE/THREONINE PROTE	1.47e+00
17	70	20.9	316	1	PP1_MAIZE	SERINE/THREONINE PROTE	1.47e+00
18	70	20.9	359	1	HIS8_SALTY	HISTIDINOL-PHOSPHATE A	1.47e+00
19	70	20.9	368	1	1AIC_PHAAU	1 AMINOCYCLOPROPANE-1-	1.47e+00
20	70	20.9	506	1	TDT_CHICK	DNA NUCLEOTIDYLEXOTRAN	1.47e+00
21	70	20.9	610	1	LEM2_HUMAN	E-SELECTIN PRECURSOR (	1.47e+00
22	70	20.9	664	1	PDI3_RAT	PROTEIN-ARGININE DEIMI	1.47e+00
23	69	20.6	312	1	PP12_YEAST	SERINE/THREONINE PROTE	2.19e+00
24	69	20.6	357	1	[illegible]_BRAJA	SERINE/THRE	[illegible]
25	69	20.6	418	1	PP11_ARATH	SERINE/THRE	[illegible]
26	69	20.6	424	1	PP1_EMENI	SERINE/THRE	[illegible]
27	69	20.6	427	1	PP11_SCHPO	SERINE/THRE	[illegible]
28	69	20.6	449	1	AFO_BRANA	ANTER-SPECI	[illegible]
29	69	20.6	487	1	HEMA_PHDBV	HEMAGGLUTIN	[illegible]
30	69	20.6	664	1	VIER_BSV7J	PROBABLE DNA	[illegible]
31	69	20.6	669	1	PDI2_HUMAN	PROTEIN ARG	[illegible]
32	68	20.3	327	1	[illegible]_EYSA	SERINE/THRE	[illegible]
33	68	20.3	427	1	MFGM_RAT	LACTADHERIN	[illegible]
34	68	20.3	1442	1	YJ9F_YEAST	HYPOTHETICA	[illegible]
35	67	20.0	177	1	KITH_MSNPV	THYMIDINE K	[illegible]
36	67	20.0	177	1	KITH_VACCV	THYMIDINE K	[illegible]
37	67	20.0	177	1	KITH_VARV	THYMIDINE K	[illegible]
38	67	20.0	193	1	[illegible]X2_MYTED	CYTOCHROME	[illegible]
39	67	20.0	492	1	PDI4_DROME	SERINE/THRE	[illegible]
40	67	20.0	518	1	KKL6_YEAST	PROBABLE SE	[illegible]
41	67	20.0	2191	1	[illegible]_XANA6	GENOME P	[illegible]
42	67	20.0	2211	1	[illegible]_MCPA	GENOME P	[illegible]
43	66	19.7	329	1	PP11_ACACL	SERINE/THRE	[illegible]
44	66	19.7	483	1	YBI1_ECOLI	HYPOTHETICA	[illegible]
45	66	19.7	1285	1	SL17_ENTHI	170 KDA SUR	[illegible]

ALIGNMENTS

RESULT 1
ID PDI3_MOUSE STANDARD; PRT; 664 AA.
AC Q9Z184;
DT 15-FEB-2000 (Rel. 39, Created)
DT 15-FEB-2000 (Rel. 39, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE PROTEIN ARGININE DEIMINASE TYPE III (EC 4.5.3.15) [illegible]
DE DEIMINASE III).
GN PDI3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EPIDERMIS;
RX MEDLINE; 99192810.
RA Rusd A.A., Ikejiri Y., Ono E., Yonekawa T., Shira [illegible]
RA Takahara H.;
RT "Molecular cloning of cDNAs of mouse peptidylargin[illegible]
RT I, type III and type IV, and the expression patte[illegible]
RT mouse.";
RL Eur. J. Biochem. 259:660-669(1999).
RN [2]
RP CHARACTERIZATION, AND TISSUE SPECIFICITY.
RX MEDLINE; 92138676.
RA Terakawa H., Takahara H., Sugawara K.;
RT "Three types of mouse peptidylarginine deiminase: [illegible]
RT tissue distribution.";
RL J. Biochem. 110:661-666(1991).
CC -!- FUNCTION: CATALYZES THE DEIMINATION OF ARGININ[illegible]
CC PROTEINS.
CC -!- CATALYTIC ACTIVITY: PROTEIN L-ARGININE + H(2)
CC L-CITRULLINE + NH(3).
CC -!- COFACTOR: REQUIRES CALCIUM IONS.
CC -!- TISSUE SPECIFICITY: EPIDERMIS AND HAIR FOLLI[illegible]
CC -!- SIMILARITY: SOME SIMILARITY WITH ARGINASE AND
CC DEIMINASE.
CC ----------[illegible]
CC This SWISS-PROT entry is copyright. It is produced [illegible]
CC between the Swiss Institute of Bioinformatics an[illegible]
CC the European Bioinformatics Institute. There are[illegible]
CC use by non-profit institutions as long as its [illegible]
CC modified and this statement is not removed. Usag[illegible]
CC entities requires a license agreement (See http:/[illegible]
CC or send an email to license@isb-sib.ch).
CC ----------[illegible]
DR EMBL; AB013849; BAA34182.1; -.

```
DR   MGD; MGI:1338891; PDB.
KW   Hydrolase; Calcium-binding; Multigene family.
FT   CA_BIND     504     515       POTENTIAL.
SQ   SEQUENCE   664 AA;  75099 MW;  BB005AE930AE6F31 CRC64;

  Query Match           23.3%;  Score 78;  DB 1;  Length 664;
  Best Local Similarity  33.3%;  Pred. No. 5.40e-02;
  Matches   10;  Conservative    8;  Mismatches  11;  Indels   1;  Gaps   1;

Db    284 IFTDIAVFRVAPWIMTPSTLPPLEVYV-CR 312
          || :::::  :  |   | ||||:  | |:
Qy     15 IFNEASIYCQGKFINGTSVLPPLDNSVLCK 44


RESULT    2
ID   SYEM_YEAST     STANDARD;      PRT:   536 AA.
AC   P48525; Q08203;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   GLUTAMYL-TRNA SYNTHETASE, MITOCHONDRIAL (EC 6.1.1.17) (GLUTAMATE--TRNA
DE   LIGASE) (GLURS).
GN   MSE1 OR YOL033W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D273-10B;
RA   Tzagoloff A.A., Shtanko A.;
RL   Submitted (JAN 1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Habbig B., Battenhorst U., Hollenberg C.P., Pamezani Rad M.;
RL   Submitted (JUL 1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: ATP + L-GLUTAMATE + TRNA(GLU) = AMP +
CC       PYROPHOSPHATE + L-GLUTAMYL-TRNA(GLU).
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC   -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L39015; AAA61403.1; -.
DR   EMBL; Z74775; CAA99033.1; -.
DR   HSSP; P27000; 1GLN.
DR   SGD; L0002617; MSE1.
DR   PFAM; PF00749; tRNA-synt_1c; 1.
DR   PRINTS; PR00987; TRNASYNTHGLU.
DR   PROSITE; PS00178; AA_TRNA_LIGASE_1; FALSE_NEG.
KW   Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW   Mitochondrion.
FT   SIMILAR      53      61       "HIGH" REGION.
FT   CONFLICT    464     464       F -> L (IN REF. 1).
SQ   SEQUENCE   536 AA;  61603 MW;  5CF36FBAD0E8C58C CRC64;

  Query Match           22.7%;  Score 76;  DB 1;  Length 536;
  Best Local Similarity  44.0%;  Pred. No. 1.26e-01;
  Matches   11;  Conservative    8;  Mismatches   4;  Indels   2;  Gaps   2;

Db     11 YCSPSKIKCVCGYSPTPKSLSES 34
          || :|:|:| : |:|| :|:| |:
Qy     22 YC-QGKFINGTSVLPPLDNSVLCKR 45


RESULT    3
ID   FKB2_METJA     STANDARD;      PRT:   231 AA.
AC   Q58235;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   PUTATIVE FKBP-TYPE PEPTIDYL-PROLYL CIS-TRANS ISOMERASE MJ0825
DE   (EC 5.2.1.8) (PPIASE) (ROTAMASE).
GN   MJ0825.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC   Methanococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE; 96337999.
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   !  FUNCTION: PPIASES ACCELERATE THE FOLDING OF PROTEINS.
CC   !  CATALYTIC ACTIVITY. CIS TRANS ISOMERISATION OF PROLINE IMIDIC
CC      PEPTIDE BONDS IN OLIGOPEPTIDES.
CC   !  SIMILARITY: BELONGS TO THE FKBP TYPE PPIASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U67526; AAB98824.1; -.
DR   HSSP; Q00688; 1PBK.
DR   TIGR; MJ0825; -.
DR   PFAM; PF00254; FKBP; 1.
DR   PROSITE; PS00453; FKBP_PPIASE_1; FALSE_NEG.
DR   PROSITE; PS00454; FKBP_PPIASE_2; 1.
DR   PROSITE; PS50059; FKBP_PPIASE_3; 1.
KW   Hypothetical protein; Isomerase; Rotamase.
FT   DOMAIN       5      92       PPIASE, FKBP-TYPE.
SQ   SEQUENCE   231 AA;  25947 MW;  ABD9AD1DB1B6C420 CRC64;

  Query Match           22.4%;  Score 75;  DB 1;  Length 231;
  Best Local Similarity  41.4%;  Pred. No. 1.92e-01;
  Matches   12;  Conservative    5;  Mismatches  11;  Indels   1;  Gaps   1;

Db     34 IYNPAMIYGPVAIFAGEGQVLPGLDEAIL 62
          |:| | ||   : |    ||| ||:::|
Qy     15 IFNEASIYCQ-GKFINGTSVLPPLDNSVL 42


RESULT    4
ID   LEM2_CANFA     STANDARD;      PRT:   611 AA.
AC   P33730;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   P-SELECTIN PEPTIDE-R (ENDOTHELIAL LEUKOCYTE ADHESION MOLECULE 1)
DE   (ELAM-1) (LEUKOCYTE-ENDOTHELIAL CELL ADHESION MOLECULE 2) (LECAM2)
DE   (CD62E).
GN   SELE.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=JUGULAR VEIN;
```

```
RA   Manning A.M., Lane C.I., Auchampach J.A., Kukielka G.L.,
RA   Rosenbloom C.L., Anderson D.C.;
RL   Submitted (XXX 1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: EXPRESSED ON CYTOKINE INDUCED ENDOTHELIAL CELLS AND
CC       MEDIATES THEIR BINDING TO LEUKOCYTES. THE LIGAND RECOGNIZED BY
CC       ELAM-1 IS SIALYL-LEWIS X (ALPHA(1->3)FUCOSYLATED DERIVATIVES OF
CC       POLYLACTOSAMINE THAT ARE FOUND AT THE NONREDUCING TERMINI OF
CC       GLYCOLIPIDS).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- SIMILARITY: TO OTHER SELECTINS/LECAMS.
CC   -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 EGF LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 6 SUSHI (SCR) REPEATS.
CC   ---------- ----------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ----------
DR   EMBL; L24087; AAA30843.1; -.
DR   HSSP; P16581; 1ESL.
DR   PFAM; PF00008; EGF; 1.
DR   PFAM; PF00059; lectin_c; 1.
DR   PFAM; PF00084; sushi; 6.
DR   PRINTS; PR00343; SELECTIN.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;
KW   Selectin; Signal; Sushi; Repeat.
FT   SIGNAL        1     22       BY SIMILARITY.
FT   CHAIN        23    611       E-SELECTIN.
FT   DOMAIN       23    557       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    558    579       POTENTIAL.
FT   DOMAIN      580    611       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       39    139       C-TYPE LECTIN (SHORT FORM).
FT   DOMAIN      140    176       EGF-LIKE.
FT   DOMAIN      180    549       6 X SUSHI (SCR) REPEATS.
FT   REPEAT      180    239       SUSHI 1.
FT   REPEAT      242    301       SUSHI 2.
FT   REPEAT      304    364       SUSHI 3.
FT   REPEAT      367    427       SUSHI 4.
FT   REPEAT      430    490       SUSHI 5.
FT   REPEAT      493    549       SUSHI 6.
FT   DISULFID     41    139       BY SIMILARITY.
FT   DISULFID    112    131       BY SIMILARITY.
FT   DISULFID    144    155       BY SIMILARITY.
FT   DISULFID    149    164       BY SIMILARITY.
FT   DISULFID    166    175       BY SIMILARITY.
FT   DISULFID    181    225       BY SIMILARITY.
FT   DISULFID    211    238       BY SIMILARITY.
FT   DISULFID    243    287       BY SIMILARITY.
FT   DISULFID    273    300       BY SIMILARITY.
FT   DISULFID    305    350       BY SIMILARITY.
FT   DISULFID    336    363       BY SIMILARITY.
FT   DISULFID    368    413       BY SIMILARITY.
FT   DISULFID    399    426       BY SIMILARITY.
FT   DISULFID    431    476       BY SIMILARITY.
FT   DISULFID    462    489       BY SIMILARITY.
FT   DISULFID    494    535       BY SIMILARITY.
FT   DISULFID    521    548       BY SIMILARITY.
FT   CARBOHYD     26     26       POTENTIAL.
FT   CARBOHYD    161    161       POTENTIAL.
FT   CARBOHYD    204    204       POTENTIAL.
FT   CARBOHYD    266    266       POTENTIAL.
FT   CARBOHYD    313    313       POTENTIAL.
FT   CARBOHYD    333    333       POTENTIAL.
FT   CARBOHYD    528    528       POTENTIAL.
SQ   SEQUENCE   611 AA;  66314 MW;  35DA9E3DF225E4F6 CRC64;
```

```
  Query Match          22.4%;  Score 75;  DB 1;  Leng
  Best Local Similarity  [illegible];  Pred. No. 1.92e-01;
  Matches    7;  Conservative    4;  Mismatches   [illegible]

Db      31 AMTFDEASTYCQQRYTHLVAI 51
           :: |:||| ||| :: : ::
Qy      13 SIIFNEASIYCQGKFTNGISV 33


RESULT    5
ID   SRE2_CAEEL     STANDARD;      PRT;   616 AA.
AC   Q09273;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   SRE-2 PROTEIN.
GN   SRE-2 OR C41C4.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditi[illegible]
OC   Rhabditidae; Peloderinae; Caenorhabditis
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Burton J.;
RL   Submitted (JAN-1995) to the EMBL/GenBank/DDBJ dat[illegible]
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTE[illegible]
CC   -!- SIMILARITY: BELONGS TO [illegible] C.ELEGANS RECEPTOR [illegible]
CC       FAMILY.
CC   ---------- ---------- [illegible] ----------
CC   This SWISS-PROT entry is [illegible]opyright. It is produced [illegible]
CC   between the Swiss Institute of Bioinformatics [illegible]
CC   the European Bioinformatics Institute. There are [illegible]
CC   use by non-profit institutions as long as [illegible]
CC   modified and this statement is not removed. Usage [illegible]
CC   entities requires a license agreement (See http:/[illegible]
CC   or send an email to license@isb-sib.ch).
CC   ---------- ---------- [illegible] ----------
DR   EMBL; Z48045; CAA88099.1; -.
DR   WORMPEP; C41C4.2; CE01517.
KW   Transmembrane; Multigene family.
FT   TRANSMEM      7     27       POTENTIAL.
FT   TRANSMEM    129    149       POTENTIAL.
FT   TRANSMEM    179    199       POTENTIAL.
FT   TRANSMEM    269    289       POTENTIAL.
FT   TRANSMEM    303    323       POTENTIAL.
FT   TRANSMEM    396    416       POTENTIAL.
FT   TRANSMEM    426    446       POTENTIAL.
FT   TRANSMEM    482    502       POTENTIAL.
FT   TRANSMEM    522    542       POTENTIAL.
SQ   SEQUENCE   616 AA;  [illegible] MW;  [illegible]CCAACE C

  Query Match          22.4%;  Score 75;  DB 1;  Len
  Best Local Similarity  [illegible];  Pred. No. 1.92e-01;
  Matches   11;  Conservative    5;  Mismatches  12;  [illegible]

Db     509 LKNDYPVLVKCCHIELNIGIAIYAQVVFF 537
           : | |||   || :::   :|' | |:
Qy       1 MINKYIVKSICCSIIN EAIIY QKFI 28


RESULT    6
ID   RRPB_CVMA5     STANDARD;      PRT;  2733 AA.
AC   P16342;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (GKFIP
GN   P2.
OS   Murine coronavirus MHV (strain A59).
OC   Viruses; ssRNA positive-strand viruses, no DNA s[illegible]
OC   Coronaviridae; Coronavirus.
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90245573.
RA   Bredenbeek P.J., Pachuk C.J., Noten A.F.H., Charite J., Luytjes W.,
RA   Weiss S.R., Spaan W.J.M.;
RT   "The primary structure and expression of the second open reading
RT   frame of the polymerase gene of the coronavirus MHV-A59; a highly
RT   conserved polymerase is expressed by an efficient ribosomal
RT   frameshifting mechanism.";
RL   Nucleic Acids Res. 18:1825-1832(1990).
CC   -!- FUNCTION: THE RNA DEPENDENT RNA POLYMERASE OF CORONAVIRUSES IS
CC       A MULTIFUNCTIONAL PROTEIN: IT CONTAINS THE ACTIVITIES NECESSARY
CC       FOR THE TRANSCRIPTION OF NEGATIVE STRANDED RNA, LEADER RNA,
CC       SUBGENOMIC MRNAS AND PROGENY VIRION RNA.
CC   -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE + N PYROPHOSPHATE
CC       + RNA(N).
CC   -!- MISCELLANEOUS: THIS PROTEIN MAY BE TRANSLATED AS A 1A-1B
CC       POLYPROTEIN BY A RIBOSOMAL FRAMESHIFTING MECHANISM.
CC   -!- SIMILARITY: TO THE AVIAN CORONAVIRUS IBV RNA POLYMERASE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X51939; CAA36202.1; ALT_SEQ.
DR   PIR; S08652; S08652.
DR   PIR; S15760; S15760.
KW   Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT   DOMAIN      622    824       POLYMERASE.
FT   DOMAIN      944   1014       CYS/HIS-RICH.
FT   DOMAIN     1218   1506       HELICASE.
FT   NP_BIND    1220   1227       ATP (BY SIMILARITY).
SQ   SEQUENCE   2733 AA;  309218 MW;  F3ACC8EF20D20C41 CRC64;

  Query Match           22.4%;  Score 75;  DB 1;  Length 2733;
  Best Local Similarity  26.2%;  Pred. No. 1.92e-01;
  Matches   11;  Conservative   14;  Mismatches  14;  Indels   3;  Gaps   3;

Db    908 DSYSVILSTCIGQKFTIETFY-KNMYLR-SAVLQSVCACVVC 947
          : |:|  ||| :  | : :::   :::  :::| ::   |:|
Qy      3 NKYPVK-STCCSTIFNEASYCGGKFINGTSVLPPLDNSVLC 43


RESULT    7
ID   KMHC_DICDI     STANDARD;      PRT:   784 AA.
AC   P34125;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   MYOSIN HEAVY CHAIN KINASE (EC 2.7.1.129) (MHCK).
GN   MHCK.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Dictyostelida; Dictyostelium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX3;
RX   MEDLINE; 92335200.
RA   Ravid S., Spudich J.A.;
RT   "Membrane-bound Dictyostelium myosin heavy chain kinase: a
RT   developmentally regulated substrate specific member of the protein
RT   kinase C family.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:5877-5881(1992).
CC   -!- FUNCTION: BRINGS ABOUT PHOSPHORYLATION OF SPECIFIC THREONINES IN
CC       THE HEAVY CHAINS OF MYOSIN DURING CHEMOTAXIS, WHICH RESULTS IN
CC       INHIBITION OF MYOSIN THICK FILAMENT FORMATION. PHOSPHORYLATION
CC       ALSO INHIBITS THE ACTIN-ACTIVATED ATPASE ACTIVITY OF MYOSIN.
CC   -!- CATALYTIC ACTIVITY: ATP + [MYOSIN HEAVY-CHAIN] = ADP +
CC       [MYOSIN HEAVY-CHAIN] PHOSPHATE.
CC   -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED ONLY DURING DEVELOPMENT.
CC   -!- SIMILARITY: BELONGS TO THE EUKARYOTIC DIACYLGLYCEROL KINASE
CC       FAMILY.
CC   -!- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL ESTER AND DAG
CC       BINDING DOMAINS.
CC   -!- CAUTION: WHILE THIS IS EXPERIMENTALLY DEFINED TO BE A PROTEIN
CC       KINASE, IT COULD BE, ON THE BASIS OF SEQUENCE SIMILARITIES, A
CC       DIACYLGLYCEROL KINASE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M93393; -; NOT_ANNOTATED_CDS.
DR   PIR; A46136; A46136.
DR   DICTYDB; DD01046; MHCK.
DR   PFAM; PF00609; DAGKa; 1.
DR   PFAM; PF00781; DAGKc; 1.
DR   PFAM; PF00130; DAG_PE-bind; 2.
DR   PROSITE; PS00479; DAG_PE_BINDING_DOMAIN; 1.
KW   Transferase; Serine/threonine protein kinase; ATP-binding; Membrane;
KW   Phosphorylation; Phorbol-ester binding.
FT   DOMAIN       19     69       PHORBOL ESTER AND DAG BINDING
FT                                (DEGENERATE).
FT   DOMAIN       88    138       PHORBOL ESTER AND DAG BINDING
FT                                (DEGENERATE).
FT   NP_BIND     466    474       ATP (BY SIMILARITY).
FT   BINDING     489    489       ATP (BY SIMILARITY).
FT   DOMAIN      188    196       ASN-RICH.
FT   DOMAIN      754    773       SER/THR-RICH.
FT   MOD_RES     768    768       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ   SEQUENCE   784 AA;  89154 MW;  6CAF177B443BF43E CRC64;

  Query Match           22.1%;  Score 74;  DB 1;  Length 784;
  Best Local Similarity  60.0%;  Pred. No. 2.91e-01;
  Matches   12;  Conservative    2;  Mismatches   5;  Indels   1;  Gaps   1;

Db    487 QTKEIVGVTVLPHLDD-VLC 505
          | | | | :||| ||: |||
Qy     24 QGKFINGTSVLPPLDNSVLC 43


RESULT    8
ID   ENV_SIVAT      STANDARD;      PRT:   865 AA.
AC   P05886;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-NOV-1988 (Rel. 09, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   ENVELOPE POLYPROTEIN GP160 PRECURSOR [CONTAINS: EXTERIOR MEMBRANE
DE   GLYCOPROTEIN (GP120); TRANSMEMBRANE GLYCOPROTEIN (GP41)].
GN   ENV.
OS   Simian immunodeficiency virus (TYO-1 isolate) (SIV-AGM).
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88232906.
RA   Fukasawa M., Miura T., Hasegawa A., Morikawa S., Tsujimoto H.,
RA   Miki K., Kitamura T., Hayami M.;
RT   "Sequence of simian immunodeficiency virus from African green monkey,
RT   a new member of the HIV/SIV group.";
RL   Nature 333:457-461(1988).
CC   -!- MISCELLANEOUS: THIS IS AN AFRICAN GREEN MONKEY ISOLATE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
CC   ---------- ----------
DR   EMBL; X07805; CAA30663.2; -.
DR   PIR; G30045; VCLJC4.
DR   HIV; X07805; ENVSACMIY.
DR   PFAM; PF00516; GP120; 1.
DR   PFAM; PF00517; GP41; 1.
KW   AIDS; Coat protein; Polyprotein; Glycoprotein; Transmembrane;
KW   Signal.
FT   SIGNAL        1    718
FT   CHAIN       719    536       EXTERIOR MEMBRANE GLYCOPROTEIN.
FT   CHAIN       537    865       TRANSMEMBRANE GLYCOPROTEIN.
FT   SITE        770    770       IN-FRAME TERMINATION CODON.
FT   CARBOHYD     35     35       POTENTIAL.
FT   CARBOHYD     68     68       POTENTIAL.
FT   CARBOHYD    117    117       POTENTIAL.
FT   CARBOHYD    150    150       POTENTIAL.
FT   CARBOHYD    165    165       POTENTIAL.
FT   CARBOHYD    195    195       POTENTIAL.
FT   CARBOHYD    198    198       POTENTIAL.
FT   CARBOHYD    210    210       POTENTIAL.
FT   CARBOHYD    252    252       POTENTIAL.
FT   CARBOHYD    255    255       POTENTIAL.
FT   CARBOHYD    266    266       POTENTIAL.
FT   CARBOHYD    276    276       POTENTIAL.
FT   CARBOHYD    282    282       POTENTIAL.
FT   CARBOHYD    294    294       POTENTIAL.
FT   CARBOHYD    306    306       POTENTIAL.
FT   CARBOHYD    316    316       POTENTIAL.
FT   CARBOHYD    373    373       POTENTIAL.
FT   CARBOHYD    414    414       POTENTIAL.
FT   CARBOHYD    451    451       POTENTIAL.
FT   CARBOHYD    488    488       POTENTIAL.
FT   CARBOHYD    491    491       POTENTIAL.
FT   CARBOHYD    645    645       POTENTIAL.
FT   CARBOHYD    661    661       POTENTIAL.
SQ   SEQUENCE   865 AA;  99024 MW;  6CEF0F0900ID6D95 CRC64;

  Query Match           22.1%;  Score 74;  DB 1;  Length 865;
  Best Local Similarity  44.0%;  Pred. No. 2.9le-01;
  Matches   11;  Conservative    4;  Mismatches   9;  Indels   1;  Gaps   1;

Db    178 KKNYYSVVWNDAEIYCKNK-TNSTS 201
          |    |:: |:| ||| | |:||
Qy      8 KSTCCSIIPNEASIYCQCPPINGIS 32


RESULT    9
ID   YIP3_YEAST     STANDARD;      PRT:   393 AA.
AC   P40454;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-FEB-1995 (Rel. 31, Last annotation update)
DE   HYPOTHETICAL 45.1 KDA PROTEIN IN IMP2 DNA13 INTERGENIC REGION.
GN   YIL153W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA   Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA   Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA   Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA   Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA   Walsh S.V., Whitehead S.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
CC   ---------- ----------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://
CC   or send an email to license@isb-sib.ch).
CC   ---------- ----------
DR   EMBL; X38059; CAA86125.1; -.
DR   PIR; S48381; S48381.
KW   Hypothetical protein.
SQ   SEQUENCE   393 AA;  45 82 MW;  22F52B43597?9BCE CF

  Query Match           21.8%;  Score 74;  DB 1;  Length
  Best Local Similarity  ?1.?%;  Pred. No. 4.40e-01;
  Matches    8;  Conservative    11;  Mismatches    4;

Db    242 YKDTNPYCQGINFTNEVK MSTCE 269
          ::::::|||| :|||   : ::::
Qy     16 PNEASIYCQC KFTNGINVTFTLN 39


RESULT   10
ID   LEM2_PIG       STANDARD;      PRT:   484 AA.
AC   P98110;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   E SELECTIN PRECURSOR (ENDOTHELIAL LEUKOCYTE ADHES
DE   (ELAM 1) (LEUKOCYTE ENDOTHELIAL CELL ADHESION MOLE
DE   (CD62E).
GN   SELE.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrat
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suida
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE AORTIC ENDOTHELIUM;
RX   MEDLINE; 95071392.
RA   Rollins S.A., Evans M.J., Johnson K.K., Elliot E.
RA   Matis L.A., Rother R.P.;
RT   "Molecular and functional analysis of porcine E-se
RT   potential role in xenograft rejection.";
RL   Biochem. Biophys. Res. Commun. 204:763-771(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=AORTIC ENDOTHELIUM;
RX   MEDLINE; 94271236.
RA   Tsang Y.T.M., Haskard D., Robinson M.K.;
RT   "Cloning and expression kinetics of porcine vascu
RT   molecule.";
RL   Biochem. Biophys. Res. Commun. 201:805-805(1994).
CC   -!- FUNCTION: EXPRESSION CYTOKINE INDUCED END
CC       MEDIATES THEIR BINDING TO LEUKOCYTES. THE LIGA
CC       ELAM-1 IS SIALYL LEWIS X (ALPHA(1-3)FUCOSYLA
CC       POLYLACTOSAMINE THAT ARE FOUND AT THE N-LINKED
CC       GLYCOLIPIDS).
CC   -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN ACUTE I
CC       REJECTION AND PROBABLY ALSO IN XENOGRAFT REJE
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN
CC   -!- SIMILARITY: TO OTHER SELECTINS/LECAMS.
CC   -!- SIMILARITY: CONTAINS 1 C TYPE LECTIN FAMILY D
CC   -!- SIMILARITY: CONTAINS 1 EGF LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 4 SUSHI (SCR) REPEATS; I
CC       THE HUMAN SUSHI 1 AND 4 EQUIVALENTS.
CC   ---------- ----------
CC   This SWISS-PROT entry is copyright. It is produce
CC   between the Swiss Institute of Bioinformatics a
CC   the European Bioinformatics Institute. There are
CC   use by non-profit institutions as long as its
CC   modified and this statement is not removed. Usage
CC   entities requires a license agreement (See http:/
CC   or send an email to license-ist@isb-sib.ch).
CC   ---------- ----------
DR   EMBL; L39076; AAA61945.1; -.
DR   EMBL; U08350; AAA21541.1; -.
DR   HSSP; P16581; 1PSL.
DR   PFAM; PF00008; EGF; 1.
```

```
DR   PFAM; PF00059; lectin_c; 1.
DR   PFAM; PF00084; sushi; 4.
DR   PRINTS; PR00343; SELECTIN.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;
KW   Selectin; Signal; Sushi; Repeat.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    484       E-SELECTIN.
FT   DOMAIN       23    429       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    430    451       POTENTIAL.
FT   DOMAIN      452    484       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       39    139       C-TYPE LECTIN (SHORT FORM)
FT   DOMAIN      140    176       EGF-LIKE.
FT   DOMAIN      180    421       4 X SUSHI (SCR) REPEATS.
FT   REPEAT      180    236       SUSHI 1.
FT   REPEAT      239    299       SUSHI 2.
FT   REPEAT      302    362       SUSHI 3.
FT   REPEAT      365    421       SUSHI 4.
FT   DISULFID     41    139       BY SIMILARITY.
FT   DISULFID    112    131       BY SIMILARITY.
FT   DISULFID    144    155       BY SIMILARITY.
FT   DISULFID    149    164       BY SIMILARITY.
FT   DISULFID    166    175       BY SIMILARITY.
FT   DISULFID    181    222       BY SIMILARITY
FT   DISULFID    208    245       BY SIMILARITY.
FT   DISULFID    240    285       BY SIMILARITY.
FT   DISULFID    271    298       BY SIMILARITY.
FT   DISULFID    303    348       BY SIMILARITY.
FT   DISULFID    334    361       BY SIMILARITY.
FT   DISULFID    366    407       BY SIMILARITY.
FT   DISULFID    393    420       BY SIMILARITY.
FT   CARBOHYD     61     61       POTENTIAL.
FT   CARBOHYD     65     65       POTENTIAL.
FT   CARBOHYD     79     79       POTENTIAL.
FT   CARBOHYD    160    160       POTENTIAL.
FT   CARBOHYD    201    201       POTENTIAL.
FT   CARBOHYD    254    254       POTENTIAL.
FT   CARBOHYD    376    376       POTENTIAL.
FT   CARBOHYD    400    400       POTENTIAL.
FT   CONFLICT    253    253       C -> Y (IN REF. 2).
FT   CONFLICT    313    313       L -> F (IN REF. 2).
FT   CONFLICT    321    321       T -> N (IN REF. 2).
FT   CONFLICT    327    327       K -> N (IN REF. 2).
FT   CONFLICT    363    363       V -> A (IN REF. 2).
FT   CONFLICT    384    384       V -> M (IN REF. 2).
FT   CONFLICT    461    484       KEVPSSSSECLQPNGSYQMPSDLI ->
FT                                NLFLPAAPNAFNPMDPIKCLLI (IN REF. 2).
SQ   SEQUENCE   484 AA;  52567 MW;  AFF74FE25C1ED013 CRC64;

  Query Match           21.8%;  Score 73;  DB 1;  Length 484;
  Best Local Similarity  28.6%;  Pred. No. 4.40e-01;
  Matches    6;  Conservative    9;  Mismatches   6;  Indels   0;  Gaps   0;

Db      31 TMIFDDASAYCQQRYTHLVAI 51
           :: ::::| ||| :: : ::
Qy      13 SIIFNEASIYCQGKFINGTSV 33


RESULT   11
ID   IF2_BORBU      STANDARD;      PRT;   882 AA.
AC   O51741;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   TRANSLATION INITIATION FACTOR IF-2.
GN   INFB OR BB0801.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 35210 / B31;
RX   MEDLINE; 98065943.
RA   Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA   Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA   Dougherty B., Tomb J.F., Fleischmann R.D., Richardson D.,
RA   Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA   van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA   Utterback T., Watthey L., McDonald L., Artiach P., Bowman C.,
RA   Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA   Smith H.O., Venter J.C.;
RT   "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT   burgdorferi.";
RL   Nature 390:580-586(1997).
CC   -!- FUNCTION: IF-2, ONE OF THE ESSENTIAL COMPONENTS FOR THE INITIATION
CC       OF PROTEIN SYNTHESIS IN VITRO, PROTECTS FORMYLMETHIONYL-TRNA FROM
CC       SPONTANEOUS HYDROLYSIS AND PROMOTES ITS BINDING TO THE 30S
CC       RIBOSOMAL SUBUNITS. IT IS ALSO INVOLVED IN THE HYDROLYSIS OF GTP
CC       DURING THE FORMATION OF THE 70S RIBOSOMAL COMPLEX (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- SIMILARITY: BELONGS TO THE IF-2 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE001179; AAC67153.1; -.
DR   TIGR; BB0801; -.
DR   PFAM; PF00009; GTP_EFTU; 1.
DR   PROSITE; PS01176; IF2; FALSE_NEG.
KW   Initiation factor; Protein biosynthesis; GTP-binding.
FT   DOMAIN      383    531       G-DOMAIN.
FT   NP_BIND     389    396       GTP (BY SIMILARITY).
FT   NP_BIND     435    439       GTP (BY SIMILARITY).
FT   NP_BIND     489    492       GTP (BY SIMILARITY).
SQ   SEQUENCE   882 AA,  97795 MW,  0FA20D65B5443C41 CRC64;

  Query Match           21.8%;  Score 73;  DB 1;  Length 882;
  Best Local Similarity  25.0%;  Pred. No. 4.40e-01;
  Matches    6;  Conservative   11;  Mismatches   7;  Indels   0;  Gaps   0;

Db     569 KGIVCSVIIEDGTLYVGDSFVGGA 592
           ::  ||:|:::::::|    |: |:
Qy       8 KSTCCSIIFNEASIYCQGKFINGT 31


RESULT   12
ID   MFGM_MOUSE     STANDARD;      PRT;   463 AA.
AC   P21956; P97800;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   LACTADHERIN PRECURSOR (MILK FAT GLOBULE-EGF FACTOR 8) (MFG-E8) (MFGM)
DE   (SPERM SURFACE PROTEIN SP47) (MP47).
GN   MFGE8.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 23-35.
RC   TISSUE=MAMMARY GLAND;
RX   MEDLINE; 91046008.
RA   Stubbs J.D., Lekutis C., Singer K.L., Bui A., Yuzuki D.,
RA   Srinivasan U., Parry G.;
RT   "cDNA cloning of a mouse mammary epithelial cell surface protein
RT   reveals the existence of epidermal growth factor-like domains linked
RT   to factor VIII-like sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:8417-8421(1990).
RN   [2]
RP   SEQUENCE OF 23-463 FROM N.A.
```

```
RC   TISSUE=TESTIS;
RA   Ensslin M.A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. ZONA PELLUCIDA
CC       BINDING PROTEIN.
CC   -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND
CC       SPERMATOZOAN.
CC   -!- DEVELOPMENTAL STAGE: MRNA EXPRESSION IS DETECTABLE IN MAMMARY
CC       TISSUE FROM NONPREGNANT ANIMALS & MAXIMAL IN THE LACTATING GLAND.
CC   -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC   ---------------------- ---------------------- ----------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------- ---------------------- ----------------------------
DR   EMBL; M38337; AAA39534.1; -.
DR   EMBL; Y11684; CAA72480.1; -.
DR   PIR; A36479; A36479.
DR   HSSP; P00740; 1FXA.
DR   MGD; MGI:102768; MFGE8.
DR   PFAM; PF00008; EGF; 2.
DR   PFAM; PF00754; F5_F8_type_C; 2.
DR   PRINTS; PR00010; EGFBLOOD.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
KW   Signal; Glycoprotein; Repeat; EGF-like domain; Milk.
FT   SIGNAL       1     22
FT   CHAIN       23    463       LACTADHERIN.
FT   DOMAIN      24     61       EGF-LIKE 1.
FT   DOMAIN      64    108       EGF-LIKE 2.
FT   DOMAIN     148    303       F5/8 TYPE C 1.
FT   DOMAIN     308    463       F5/8 TYPE C 2.
FT   SITE        87     89       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID    28     39       BY SIMILARITY.
FT   DISULFID    33     49       BY SIMILARITY.
FT   DISULFID    51     60       BY SIMILARITY.
FT   DISULFID    68     79       BY SIMILARITY.
FT   DISULFID    73     96       BY SIMILARITY.
FT   DISULFID    98    107       BY SIMILARITY.
FT   DISULFID   148    303       BY SIMILARITY.
FT   DISULFID   290    294       BY SIMILARITY.
FT   DISULFID   308    463       BY SIMILARITY.
FT   CARBOHYD    61     61       POTENTIAL.
FT   CARBOHYD   266    266       POTENTIAL.
FT   CARBOHYD   316    316       POTENTIAL.
FT   CARBOHYD   426    426       POTENTIAL.
FT   CONFLICT    30     30       S -> F (IN REF. 2).
FT   CONFLICT    35     35       N -> D (IN AA SEQUENCE; REF. 1).
FT   CONFLICT   110    147       ETNYYNLDGEYMPTTAVPNTAVPTPAPTPDLSNNLASR ->
FT                               G (IN REF. 2).
FT   CONFLICT   168    168       Y -> S (IN REF. 2).
FT   CONFLICT   196    196       H -> P (IN REF. 2).
FT   CONFLICT   309    309       L -> S (IN REF. 2).
FT   CONFLICT   395    395       E -> A (IN REF. 2).
SQ   SEQUENCE   463 AA;  51465 MW;  D78B6C6EFBBA724D CRC64;

  Query Match           21.5%;  Score 72;  DB 1;  Length 463;
  Best Local Similarity  29.4%;  Pred. No. 6.6le-01;
  Matches   10;  Conservative   10;  Mismatches  14;  Indels   1;  Gaps   1;

Db      16 CASGLEAASGDFCDSSICNGGTCITGQDNDIYC 49
           | | :|  :: :|:: : :|| : |   || : |
Qy      11 CGSTIPNEASIYCQGKP INGTSVLPPLDNSVLC 43
```

```
RESULT   13
ID   SP26_CAEEL     STANDARD;      PRT;   570 AA.
AC   Q10579;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   SPERMATOCYTE PROTEIN SPE-26.
GN   SPE-26 OR R10H10.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdit
OC   Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A., AND MUTAGENESIS.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 95262900.
RA   Varkey J.P., Muhlrad P.J., Minniti A.N., Do B., Wa
RT   "The Caenorhabditis elegans spe-26 gene is necessa
RT   spermatids and encodes a protein similar to the act
RT   proteins kelch and seipin.";
RL   Genes Dev. 9:1074-1086(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Bardwell S.;
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ data
CC   -!- FUNCTION: MAY PLAY A ROLE IN THE SPERMATOCYTE
CC       POSSIBLY INTERACTING WITH ACTIN.
CC   -!- TISSUE SPECIFICITY: TESTIS, IN BOTH SPERMATOCY
CC       SPERMATOCYTES.
CC   -!- SIMILARITY: BELONGS TO THE KLHDC9/MIPP FAMILY.
CC   ---------------------- ---------- ----------
CC   This SWISS-PROT entry is copyright. It is produced
CC   between the Swiss Institute of Bioinformatics and
CC   the European Bioinformatics Institute. There are
CC   use by non-profit institutions as long as its
CC   modified and this statement is not removed. Usage
CC   entities requires a license agreement (See http://
CC   or send an email to license@isb-sib.ch).
CC   ---------------------- ---------- ----------
DR   EMBL; U20142; AAB60245.1; -.
DR   EMBL; Z70686; CAA94610.1; -.
DR   WORMPEP; R10H10.2; CE06203.
DR   PFAM; PF01344; Kelch; 1.
DR   PRINTS; PR00501; KELCHREPEAT.
KW   Cytoskeleton; Repeat.
FT   DOMAIN      244    460       6 X APPROXIMATE TANDE
FT   REPEAT      244    291       1.
FT   REPEAT      292    341       2.
FT   REPEAT      341    393       3.
FT   REPEAT      394    440       4.
FT   REPEAT      441    486       5.
FT   MUTAGEN     360    360       S -> N: IN ITS1-2; TA
FT                                25 DEGREES CELSIUS.
FT   MUTAGEN     446    446       T -> I: IN HC14-2; TAL
FT                                25 DEGREES CELSIUS.
SQ   SEQUENCE   570 AA;  [illegible] MW;  B2092B2B79380B67 CR

  Query Match           21.5%;  Score 72;  DB 1;  Length
  Best Local Similarity  44.4%;  Pred. No. 6.6le-01;
  Matches   11;  Conservative    9;  Mismatches  16;

Db     382 MKSDCAAEVF-ENGTTVSGSL NGDMILASVE 411
           ::| | : :| | |  |    ' :  :::::
Qy       7 VKSTCGSIIPNEASIY'QKEINTTSVLPPLD 38


RESULT   14
ID   VC04_SPVKA     STANDARD;      PRT;   530 AA.
AC   P42228;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   PROTEIN C4.
```

```
GN   C4L.
OS   Swinepox virus (strain Kasza) (SPV).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Suipoxvirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94069924.
RA   Massung R.F., Jayarama V., Moyer R.W.;
PT   "DNA sequence analysis of conserved and unique regions of swinepox
RT   virus: identification of genetic elements supporting phenotypic
RT   observations including a novel G protein-coupled receptor
RT   homologue.";
RL   Virology 197:511-528(1993).
CC   -!- SIMILARITY: TO 16, 18 AND 19 PROTEINS OF SHOPE FIBROMA VIRUS,
CC       AND SOME SIMILARITY TO VACCINIA VIRUS C2.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L22013; AAC37867.1; -.
DR   PFAM; PF00651; BTB; 1.
DR   PFAM; PF01344; Kelch; 5.
KW   Hypothetical protein.
SQ   SEQUENCE   530 AA;  62639 MW;  31897A2422263F11 CRC64;

  Query Match           21.2%;  Score 71;  DB 1;  Length 530;
  Best Local Similarity  37.1%;  Pred. No. 9.89e-01;
  Matches   13;  Conservative    7;  Mismatches  12;  Indels   3;  Gaps   3;

Db    415 MKLKQFPIAKSGISSIVYND-FIYCIGGRIDTPHI 448
          |  | :|: ||   ||:::|:  ||| |  |:   :
Qy      1 MTNK-YPV-KSTCCSIIFNEASIYCQGKFINGTSV 33


RESULT   15
ID   POLR_EPMV      STANDARD;      PRT;  1839 AA.
AC   P20126;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01 FEB-1996 (Rel. 33, Last sequence update)
DT   15 DEC 1998 (Rel. 37, Last annotation update)
DE   RNA REPLICASE POLYPROTEIN (EC 2.7.7.48).
OS   Eggplant mosaic virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Tymovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90021185.
RA   Osorio-Keese M.E., Keese P., Gibbs A.;
RT   "Nucleotide sequence of the genome of eggplant mosaic tymovirus.";
RL   Virology 172:547-554(1989).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J04374; AAA43049.1; -.
DR   PIR; JQ0102; RRWPEM.
DR   PFAM; PF01443; Viral_helicase1; 1.
KW   Transferase; RNA-directed RNA polymerase; Polyprotein; ATP-binding.
FT   NP_BIND     965    972       ATP (BY SIMILARITY).
SQ   SEQUENCE   1839 AA;  204731 MW;  FD8DC1F5115E7861 CRC64;

  Query Match           21.2%;  Score 71;  DB 1;  Length 1839;
  Best Local Similarity  46.4%;  Pred. No. 9.89e-01;
  Matches   13;  Conservative    6;  Mismatches   8;  Indels   1;  Gaps   1;

Db   1648 YNLAVIYSQYS-LNGHPILISCDDSVLC 1664
          :| | || |   :|| ::| : |:|||
Qy     16 LNLAsIYCQGKFINGTSVLFFLDNsVLC 43


Search completed: Fri Aug 25 17:45:25 2000
Job time : 25 secs.
```

MPSRCH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 25 17:45:43 2000 MasPar time 7.94 Seconds
392.835 Million cell updates/sec

Tabular output not generated.

Title: >US-09-257-585-2
Description: (1-45) from US09257585.pep
Perfect Score: 335
Sequence: 1 MINKYDVVESIDGSIIENEASIVQQKFINCISVLPPLDNSVCKR 45

Scoring table: PAM 150
Gap 11

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl12
1:sp_archea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 32.965; Variance 47.601; scale 0.693

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	89	26.6	352	5	Q20118	F37B12.3 PROTEIN	1.05e-03
2	78	23.3	365	10	Q42917	1 AMINOCYCLOPROPANE-CA	1.29e-01
3	78	23.3	401	5	Q27495	F23B12.1 PROTEIN.	1.29e-01
4	78	23.3	421	5	Q18957	D1081.4 PROTEIN	1.29e-01
5	78	23.3	664	11	Q92184	PEPTIDYLARGININE DEIMI	1.29e-01
6	76	22.7	1048	10	Q42884	CALCIUM TRANSPORTING A	2.97e-01
7	76	22.7	2193	14	Q9WPJ0	POLYPROTEIN.	2.97e-01
8	76	22.7	2193	14	Q9WQJ0	POLYPROTEIN.	2.97e-01
9	75	22.4	146	2	Q48519	EIIA MAN.	4.49e-01
10	75	22.4	466	10	Q82218	PUTATIVE POLYGALACTURO	4.49e-01
11	75	22.4	2733	11	Q39226	RNA DIRECTED RNA POLYM	4.49e-01
12	73	21.8	153	14	Q11434	HYPOTHETICAL 16.3 KD P	1.01e+00
13	73	21.8	482	6	Q28982	E-SELECTIN.	1.01e+00
14	73	21.8	1794	14	P87515	NONSTRUCTURAL POLYPROT	1.01e+00
15	72	21.5	170	2	Q50885	CONSERVED HYPOTHETICAL	1.51e+00
16	72	21.5	426	11	Q9WTS3	MILK FAT GLOBULE GLYCO	1.51e+00
17	72	21.5	429	3	Q14135	HYPOTHETICAL 49.6 KD P	1.51e+00
18	72	21.5	429	5	Q24689	HYPOTHETICAL 48.8 KD P	1.51e+00
19	72	21.5	458	2	Q67311	CYTOPLASMIC AXIAL FILA	1.51e+00
20	72	21.5	740	10	Q92117	BETA-GALACTOSIDASE (EC	1.51e+00
21	71	21.2	144	12	Q48[illegible]	SERINE/THRE	2.12e+00
22	71	21.2	481	5	Q18146	C05A2.1 PR	2.12e+00
23	71	21.2	482	5	Q26607	T07H3.4 PR	2.12e+00
24	71	21.2	522	5	Q22441	C.ELEGANS ZK1	2.12e+00
25	71	21.2	709	14	Q59965	PRL PROTEIN	2.12e+00
26	71	21.2	715	14	Q9WMI0	PUTATIVE PR	2.12e+00
27	71	21.2	1027	5	Q20025	HYPOTHETICA	2.12e+00
28	70	20.9	114	14	Q9YN01	THYMIDINE K	2.81e+00
29	70	20.9	115	14	Q9YN02	THYMIDINE K	2.81e+00
30	70	20.9	261	5	Q20975	F48F7.3 PR	2.81e+00
31	70	20.9	256	5	Q27435	PFC0055W PR	2.81e+00
32	70	20.9	355	14	P75276	GENE ENCODI	2.81e+00
33	70	20.9	417	12	48[illegible]	SERINE/THRE	2.81e+00
34	70	20.9	484	10	Q43858	1-AMINOCYCL	2.81e+00
35	70	20.9	508	4	P24284	PHEROMONE P	2.81e+00
36	70	20.9	636	5	Q20852	COSMID F55C	2.81e+00
37	70	20.9	664	11	P70708	PEPTIDYLARG	2.81e+00
38	70	20.9	1247	5	Q17412	B0240.2 PR	2.81e+00
39	70	20.9	3083	5	Q27278	PFC0825W PR	2.81e+00
40	69	20.6	109	4	P50962	SERINE/THRE	3.72e+00
41	69	20.6	348	5	Q09854	C09H5.5 PR	3.72e+00
42	69	20.6	667	14	P46986	H PROTEIN	3.72e+00
43	69	20.6	686	4	P49208	KIAA0994 PR	3.72e+00
44	69	20.6	874	5	P16599	F21E9.5 PR	3.72e+00
45	69	20.6	1277	14	Q9BWZ2	NEURAL CELL	3.72e+00

ALIGNMENTS

RESULT 1
ID Q20118 PRELIMINARY; PRT; 352 AA.
AC Q20118;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE F37B12.3 PROTEIN.
GN F37B12.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditi
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderita
RN [1]
RP SEQUENCE FROM N.A.
RA THOMAS K.;
RL Submitted (SEP 1995) to the EMBL/GenBank/DDBJ dat
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., Cooper J.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LAI Z.,
RA LIGHTNING J., LLOYD C., MORTIMORE B., NICE J.,
RA PARSONS J., PERRY K., RIEKEN L., ROOPRA A., SAUNDERS D.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., STEINKE J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,
RT "2.2 Mb of contiguous nucleotide sequence from ch
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL; Z54218; CAA90456.1; -.
SQ SEQUENCE 352 AA; 38852 MW; 5F6E2EDF CRC32;

Query Match 26.6%; Score 89; DB 5; Le
Best Local Similarity 29.4%; Pred. No. 1.05e-03;
Matches 10; Conservative 14; Mismatches 8; Gaps

Db 229 CRIFIKALLYCSLVNLLLAAIIVNVLLRK 260
| | | | ||||||| | |||| || ||
Qy 12 CSIIENEASIVQQKFINCISVLPPLDNSVCKR 45

RESULT 2

```
ID   Q42917      PRELIMINARY;      PRT;   365 AA
AC   Q42917;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   1-AMINOCYCLOPROPANE-CARBOXYLATE SYNTHASE (EC 4.4.1.14)
DE   (1-AMINOCYCLOPROPANE-1-CARBOXYLATE SYNTHASE) (ACC SYNTHASE)
DE   (FRAGMENT)
GN   ACS.
OS   Mangifera indica (Mango).
OC   Eukaryota; Viridiplantae; Streptophyta, Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons,
OC   core eudicots; Rosidae; eurosids II; Sapindales; Anacardiaceae;
OC   Mangifera.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KENSINGTON ROYAL; TISSUE=FRUIT;
RA   BOTELLA J.R.;
RL   Submitted (MAR-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE = 1-
CC       AMINOCYCLOPROPANE-1-CARBOXYLATE + METHYLTHIOADENOSINE.
CC   -!- COFACTOR: PYRIDOXAL-PHOSPHATE.
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO CLASS-I OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC       AMINOTRANSFERASES.
DR   EMBL; U22523; AAA84895.1; -.
DR   MENDEL; 11934; Mangin;Acs;11934.
DR   PROSITE; PS00105; AA_TRANSFER_CLASS_1; 1.
DR   PFAM; PF00155; aminotran_1; 1.
KW   Lyase; Pyridoxal phosphate.
FT   NON_TER       1      1
FT   NON_TER     365    365
SQ   SEQUENCE   365 AA;  41017 MW;  E86A1FCA CRC32;

  Query Match           24.3%;  Score 78;  DB 10;  Length 365;
  Best Local Similarity  29.6%;  Pred. No. 1.29e-01;
  Matches    8;  Conservative    9;  Mismatches   9;  Indels   1;  Gaps   1;

Db    188 CDEIYA-ATLFCPPKPVSVAEVIQEMD 213
          |  |:  |:::|  ||:: : |:  :|
Qy     12 CSIIPNEASIYCQGKFINGTSVLPPLD 38


RESULT    3
ID   Q27495      PRELIMINARY;      PRT;   401 AA.
AC   Q27495;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   F23B12.1 PROTEIN.
GN   F23B12.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda, Secernentea, Rhabditia, Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   WILD A.;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z77659; CAB01164.1; -.
DR   HSSP; P08129; 1FJM.
DR   PFAM, PF00149, STphosphatase; 1.
DR   PRINTS; PR00114; STPHPHTASE.
SQ   SEQUENCE   401 AA;  46264 MW;  CEC829B7 CRC32;

  Query Match           23.3%;  Score 78;  DB 5;  Length 401;
  Best Local Similarity  37.9%;  Pred. No. 1.29e-01;
  Matches   11;  Conservative   10;  Mismatches   5;  Indels   3;  Gaps   3;

Db    333 VFSAPS-YC-GQFDNAAAVMN-VDDKLLC 358
          :|: :| || | | |:::|:  :|: :||
Qy     15 IFNEASIYCQGKFINGTSVLPPLDNSVLC 43


RESULT    4
ID   Q18957      PRELIMINARY;      PRT;   421 AA.
AC   Q18957;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   D1081.4 PROTEIN.
GN   D1081.4.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda, Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   DOBSON R.;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z75710; CAB00025.1; -.
SQ   SEQUENCE   421 AA;  48547 MW;  DC4338A2 CRC32;

  Query Match           23.3%;  Score 78;  DB 5;  Length 421;
  Best Local Similarity  35.7%;  Pred. No. 1.29e-01;
  Matches   10;  Conservative    8;  Mismatches   9;  Indels   1;  Gaps   1;

Db    389 FNEFSIYC-VRFIQSERLVTSANTIIVC 415
          ||| |||| :|| :  :: : :  ::|
Qy     16 FNEASIYCQGKFINGTSVLPPLDNSVLC 43


RESULT    5
ID   Q92184      PRELIMINARY;      PRT;   664 AA.
AC   Q92184;
DT   01 MAY 1999 (TrEMBLrel. 10, Created)
DT   01 MAY 1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   PEPTIDYLARGININE DEIMINASE TYPE III.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EPIDERMIS;
RX   MEDLINE; 99192810.
```

```
RA   RUSO A.A., IKEUCHI Y., ONO H., YONEKAWA T., SHIRAIWA M., KAWADA A.,
RA   TAKAHARA H.;
RT   "Molecular cloning of cDNAs of mouse peptidylarginine deiminase type
RT   I, type III and type IV, and the expression pattern of type I in
RT   mouse";
RL   Eur. J. Biochem. 259:660-669(1999).
DR   EMBL; AB013849; BAA34182.1; -.
SQ   SEQUENCE   664 AA;  75099 MW;  BA8BB3C3 CRC32;

  Query Match           23.3%;  Score 78;  DB 11;  Length 664;
  Best Local Similarity  33.3%;  Pred. No. 1.29e-01;
  Matches   10;  Conservative    8;  Mismatches   11;  Indels   1;  Gaps   1;

Db    284 IFIPTAVFRVAPWIMTPSTLPPLEVYV-CR 312
          || :::::  :  |   | ||||:  | |:
Qy     15 IFNEASIYCQGKFINGTSVLPPLDNSVLCK 44


RESULT    6
ID   Q42883      PRELIMINARY;      PRT:  1048 AA.
AC   Q42883;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   CALCIUM-TRANSPORTING ATPASE (EC 3.6.1.38).
GN   LCA1.
OS   Lycopersicon esculentum (Tomato).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Asteridae; euasterids I; Solanales; Solanaceae;
OC   Solanum.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93028433.
RA   WIMMERS L.E., EWING N.N., BENNETT A.B.,
RT   "Higher plant Ca(2+)-ATPase: primary structure and regulation of mRNA
RT   abundance by salt.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:9205-9209(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   NAVARRO-AVINO J.P., BENNETT A.B.,
RT   "Alternate transcription initiation sites generate two LCA1 Ca2+-
RT   ATPase mRNA transcripts in tomato";
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; M96324; AAA34138.1; -.
DR   EMBL; AF050496; AAD11618.1; -.
DR   EMBL; AF050495; AAD11617.1; -.
DR   MENDEL; 16039; Lyces;2440;16039.
DR   MENDEL; 35991; Lyces;2440;35991.
DR   MENDEL; 35992; Lyces;2440;35992.
DR   PROSITE; PS00154; ATPASE_E1_E2; 1.
DR   PFAM; PF00122; E1-E2_ATPase; 1.
DR   PRINTS; PR00119; CATATPASE.
KW   Hydrolase; Transmembrane; Phosphorylation; ATP-binding.
FT   MOD_RES     366    366       PHOSPHORYLATION (PROBABLE).
SQ   SEQUENCE   1048 AA;  116127 MW;  22173C94 CRC32;

  Query Match           22.7%;  Score 76;  DB 10;  Length 1048;
  Best Local Similarity  28.6%;  Pred. No. 2.97e-01;
  Matches   10;  Conservative   16;  Mismatches    7;  Indels   2;  Gaps   2;

Db    416 MDANLLLMAEICAIC-NDAGVFCDGRLFKATG-LP 448
          |  :  ::: |:| |:|:::|:|:::::|. ||
Qy      1 MINKYPVKSICCSIFNEASIYCQGKFINGTSVLP 35


RESULT    7
ID   Q9WPJ0      PRELIMINARY;      PRT:  2193 AA.
AC   Q9WPJ0;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   POLYPROTEIN.
OS   Enterovirus 71.
OC   Viruses; ssRNA positive-strand viruses, no DNA sta
OC   Enterovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCKU9822;
RA   YAN J.J., WANG J.R., SU I.J.;
RT   "An outbreak of Enterovirus 71 in Taiwan.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ data
DR   EMBL; AF136379; AAD44716.1;
KW   Polyprotein.
SQ   SEQUENCE   2193 AA;  242668 MW;  B66ED806 CRC32;

  Query Match           22.7%;  Score 76;  DB 14;  Len
  Best Local Similarity  40.7%;  Pred. No. 2.97e-01;
  Matches   11;  Conservative    6;  Mismatches   10;

Db    930 MRKHYPVSFSKPSLIFVEASEYYPARY 956
          | : |||  :  |:| :.|   :::
Qy      1 MINKYPVKSICCSIFNEASIY'QGKF 27


RESULT    8
ID   Q9WQJ0      PRELIMINARY;      PRT:  2193 AA.
AC   Q9WQJ0;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   POLYPROTEIN.
OS   Enterovirus 71.
OC   Viruses; ssRNA positive-strand viruses, no DNA sta
OC   Enterovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TW/2272/98;
RA   SHIH S.-R., HO M. S., CHEN Y. C., LIN K.-H., TSAO
RA   CHEN Y.-T., WU C. N., NING H. C., CHANG P.-Y., CHA
RA   LIN T.-Y.;
RT   "Sequence Analysis of Enterovirus 71 Isolated fro
RT   in Taiwan.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ data
DR   EMBL; AF119795; AAD22045.1;
KW   Polyprotein.
FT   CHAIN         1    [illegible]  VP4.
FT   CHAIN        70    [illegible]  VP2.
FT   CHAIN       324    [illegible]  VP3.
FT   CHAIN       566    [illegible]  VP1.
FT   CHAIN       863   1012        2A.
FT   CHAIN      1013   1111        2B.
FT   CHAIN      1112   1440        2C.
FT   CHAIN      1441   1526        3A.
FT   CHAIN      1527   1548        3B.
FT   CHAIN      1549   1731        3C.
FT   CHAIN      1732   2193        3D.
SQ   SEQUENCE   2193 AA;  242469 MW;  4E9100ED CRC32;

  Query Match           22.7%;  Score 76;  DB 14;  Len
  Best Local Similarity  40.7%;  Pred. No. 2.97e-01;
  Matches   11;  Conservative    6;  Mismatches   10;

Db    930 MRKHYPVSFSKPSLIFVEASEYYPARY 956
          | : |||  :  |:| |:| |  :::
Qy      1 MINKYPVKSICCSIFNEASIYCQGKF 27


RESULT    9
ID   Q48519      PRELIMINARY;      PRT:   146 AA.
AC   Q48519;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   EIIA-MAN.
GN   MANA.
```

```
OS   Lactobacillus curvatus.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC   Lactobacillus.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   VEYRAT A., GOSALBES M., PEREZ-MARTINEZ G.;
RL   Submitted (JUN-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U28163; AAB04153.1; -.
SQ   SEQUENCE   146 AA;  16060 MW;  A71DD93B CRC32;

  Query Match           22.4%;  Score 75;  DB 2;  Length 146;
  Best Local Similarity  32.1%;  Pred. No. 4.49e-01;
  Matches    9;  Conservative    8;  Mismatches  10;  Indels    1;  Gaps    1;

Db      22 MIFGENNGYCAVPPLKGEGI-PQLQAKI 48
           :|| | : ||   |:|| :: | |:  :
Qy      14 IIFNEASIYCQGKFINGTSVLPPLDNSV 41


RESULT   10
ID   O82218      PRELIMINARY;      PRT;   466 AA.
AC   O82218;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   PUTATIVE POLYGALACTURONASE.
GN   T29E15.10.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids I; Brassicales; Brassicaceae;
OC   Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN CV. COLUMBIA;
RA   ROUNSLEY S.D., LIN X., KAUL S., SHEA T.P., FUJII C.Y., MASON T.M.,
RA   SHEN M., RONNING C.M., FRASER C.M., SOMERVILLE C.R., VENTER J.C.;
RT   "Arabidopsis thaliana chromosome II BAC T29E15 genomic sequence.";
RL   Submitted (OCT 1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC005170; AAC63679.1; -.
DR   PFAM; PF00295; Glyco_hydro_28; 1.
SQ   SEQUENCE   466 AA;  50396 MW;  25703D4A CRC32;

  Query Match           22.4%;  Score 75;  DB 10;  Length 466;
  Best Local Similarity  28.6%;  Pred. No. 4.49e-01;
  Matches    6;  Conservative   11;  Mismatches   4;  Indels    0;  Gaps    0;

Db     210 PVYCNNVIVKGVTILAPIDSP 230
           ::||   :::| ::|:|:|::
Qy      20 SIYCQGKFINGISVLPPLDNS 40


RESULT   11
ID   O39226      PRELIMINARY;      PRT;  2733 AA.
AC   O39226;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JAN-1998 (TrEMBLrel. 05, Last annotation update)
DE   RNA-DIRECTED RNA POLYMERASE.
OS   murine hepatitis virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC   Coronaviridae; Coronavirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MHV-A59;
RA   LEPARC-GOFFART I., HINGLEY S.T., CHUA M.M., JIANG X., LAVI E.,
RA   WEISS S.R.;
RL   Virology 0:0-0(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MHV-A59;
RA   WEISS S.R., LEPARC-GOFFART I., HINGLEY S.T.;
RL   Submitted (OCT 1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF029248; AAB86818.1; -.
KW   RNA directed RNA polymerase.
SQ   SEQUENCE   2733 AA;  309126 MW;  09E1B333 CRC32;

  Query Match           22.4%;  Score 75;  DB 14;  Length 2733;
  Best Local Similarity  26.2%;  Pred. No. 4.49e-01;
  Matches   11;  Conservative   14;  Mismatches  14;  Indels    3;  Gaps    3;

Db     908 DSYSVILSTCDGQKFTDETFY-KNMYLR-SAVLQSVGACVVC 947
           : |:|  ||| :  | : ::|   :::  ::|| ::   |:|
Qy       3 NKYFVK-STCCSIIFNEASIYCQGKFINGTSVLPPLDNSVLC 43


RESULT   12
ID   O11434      PRELIMINARY;      PRT;   153 AA.
AC   O11434;
DT   01 JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JAN-1998 (TrEMBLrel. 05, Last annotation update)
DE   HYPOTHETICAL 16.3 KD PROTEIN.
OS   avian adenovirus EDS.
OC   Viruses, dsDNA viruses, no RNA stage, Adenoviridae, Aviadenovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=127;
RX   MEDLINE; 98042451.
RA   HESS M., BLOECKER H., BRANDT P.;
RT   "The complete nucleotide sequence of the egg drop syndrome virus: an
RT   intermediate between mastadenoviruses and aviadenoviruses.";
RL   Virology 238:145-156(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=127;
RA   HESS M.;
RL   Submitted (NOV 1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Y09598; CAA70795.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   153 AA;  16291 MW;  DBAFBA73 CRC32;

  Query Match           21.8%;  Score 73;  DB 14;  Length 153;
  Best Local Similarity  37.0%;  Pred. No. 1.01e+00;
  Matches   10;  Conservative    9;  Mismatches   6;  Indels    2;  Gaps    1;

Db      78 YTETELYCQS--ICGSIAVPPFGNPLL 102
           . |:  :|||.  | |:  .||. |..|
Qy      16 FNEASIYCQGKFINGTSVLPPLDNSVL 42


RESULT   13
ID   Q28982      PRELIMINARY;      PRT;   482 AA.
AC   Q28982;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01 NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01 NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   E SELECTIN.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97075911.
RA   WINKLER H., BROSTJAN C., CSIZMADIA V., NATARAJAN G., ANRATHER J.,
RA   BACH F.H.;
RT   "The intron-exon structure of the porcine E-selectin encoding gene.";
RL   Gene 176:67-72(1996).
DR   EMBL; U37521; AAC48680.1; -.
DR   HSSP; P16581; 1ESL.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PFAM; PF00008; EGF; 1.
DR   PFAM; PF00059; lectin_c; 1.
DR   PFAM; PF00084; sushi; 4.
DR   PRINTS; PR00343; SELECTIN.
KW   Lectin; Glycoprotein.
```

```
SQ   SEQUENCE   482 AA;  52341 MW;  94528E97 CRC32;

  Query Match           21.8%;  Score 73;  DB 6;  Length 482;
  Best Local Similarity  28.6%;  Pred. No. 1.0le+00;
  Matches    6;  Conservative    9;  Mismatches   6;  Indels   0,  Gaps   0;

Db      31 TMTFDDASAYCQQRYTHLVAI 51
           :: |:|| ||| :: :  ::
Qy      13 SLIFNEASIYCQGKFINGISV 33


RESULT   14
ID   P87515       PRELIMINARY;      PRT;  1794 AA.
AC   P87515;
DT   01-MAY-1997 (IrEMBLrel. 03, Created)
DT   01-MAY-1997 (IrEMBLrel. 03, Last sequence update)
DT   01-NOV-1999 (IrEMBLrel. 12, Last annotation update)
DE   NONSTRUCTURAL POLYPROTEIN.
OS   Barmah forest virus.
OC   Viruses; ssRNA positive strand viruses, no DNA stage; Togaviridae;
OC   Alphavirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BH2193;
RA   LEE E., STOCKS C., LOBIGS P., HISLOP A., STRAUB J., MARSHALL I.,
RA   WEIR R., DALGARNO L.;
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases
DR   EMBL; U73745; AAB40701.1; -.
DR   PFAM; PF01661; DUF27; 1.
DR   PFAM; PF01443; Viral_helicase1; 1.
KW   Polyprotein.
SQ   SEQUENCE   1794 AA;  199985 MW;  1781F9B8 CRC32;

  Query Match           21.8%;  Score 73;  DB 14;  Length 1794;
  Best Local Similarity  26.3%;  Pred. No. 1.0le+00;
  Matches   10;  Conservative   14;  Mismatches  12;  Indels   2,  Gaps   2;

Db     775 VKHTP-NILYIDEAFSCHAGTLLATIAIVRPKQKVVLC 811
           || |  :|:: : :: |: | ::   ::: | :: |||
Qy       7 VKSICCSIIFNEASIYCQ-GKFINGISVLPPLDNSVLC 43


RESULT   15
ID   O50885       PRELIMINARY;      PRT;   170 AA.
AC   O50885;
DT   01-JUN-1998 (IrEMBLrel. 06, Created)
DT   01-JUN-1998 (IrEMBLrel. 06, Last sequence update)
DT   01-NOV-1998 (IrEMBLrel. 08, Last annotation update)
DE   CONSERVED HYPOTHETICAL PROTEIN.
GN   BB131.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OG   Plasmid lp28-4.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 35210 / B31;
RX   MEDLINE; 98065943.
RA   FRASER C.M., CASJENS S., HUANG W.M., SUTTON G.G., CLAYTON R.A.,
RA   LATHIGRA R., WHITE O., KETCHUM K.A., DODSON R., HICKEY E.K., GWINN M.,
RA   DOUGHERTY B., TOMB J.-F., FLEISCHMANN R.D., RICHARDSON D.,
RA   PETERSON J., KERLAVAGE A.R., QUACKENBUSH J., SALZBERG S., HANSON M.,
RA   VAN VUGT R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,
RA   UTTERBACK T., WATTHEY L., MCDONALD L., ARTIACH P., BOWMAN C.,
RA   GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,
RA   SMITH H.O., VENTER J.C.;
RT   "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT   burgdorferi.";
RL   Nature 390:580-586(1997).
DR   EMBL; AE000789; AAC66194.1; -.
DR   TIGR; BB131; -.
KW   Plasmid.
SQ   SEQUENCE   170 AA;  18227 MW;  182689B5 CRC32;

  Query Match           21.8%;  Score 73;  DB 2;  Length
  Best Local Similarity  [illegible];  Pred. No. 1.5le+[illegible];
  Matches    9;  Conservative    6;  Mismatches   7;

Db      67 FVQGDYIGGGAVIGS[illegible] 89
           : || :| | :|| :  ::
Qy      22 YCQGKFINGISVL-PPLDNSVLC 44


Search completed: Fri Aug 25 17:47:59 2000
Job time : 73 secs.
```



Release 4.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 25 18:06:56 2000; MasPar time 4.12 Seconds
310.144 Million cell updates/sec

Tabular output not generated.

Title: >US-09-257-585-8
Description: (1-54) from US09257585.pep
Perfect Score: 398
Sequence: 1 MVIEGYLSRNSSRVCNYHGC..... NSNQLSISMDSPVEWRRRI 54

Scoring table: PAM 150
Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq36
1:geneseqp

Statistics: Mean 26.351; Variance 96.684; scale 0.273

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	86	21.6	536	1	P70138	Sequence of a region o	4.84e+00
2	75	18.8	284	1	W81733	M. tuberculosis immuno	4.36e+01
3	75	18.8	284	1	W64366	Mycobacterium tubercul	4.36e+01
4	75	18.8	386	1	W18664	Fragmented human NF-H	4.36e+01
5	71	17.8	84	1	R21200	Human parathyroid horm	9.43e+01
6	71	17.8	84	1	R21214	Human parathyroid horm	9.43e+01
7	71	17.8	419	1	P61318	Serratia phospholipase	9.43e+01
8	71	17.8	419	1	W20728	H. pylori cytoplasmic	9.43e+01
9	71	17.8	449	1	W90188	Arabidopsis Golgi glyc	9.43e+01
10	71	17.8	488	1	W89248	Human SAD.	9.43e+01
11	71	17.8	4038	1	R74171	Aspergillus terreus tr	9.43e+01
12	70	17.6	84	1	R23358	Bovine parathyroid hor	1.14e+02
13	70	17.6	84	1	R23344	Bovine parathyroid hor	1.14e+02
14	70	17.6	506	1	W44269	Hybrid DNA protein com	1.14e+02
15	69	17.3	84	1	R23541	Bovine parathyroid hor	1.38e+02
16	69	17.3	84	1	R21203	Human parathyroid horm	1.38e+02
17	69	17.3	84	1	R23374	Bovine parathyroid hor	1.38e+02
18	69	17.3	113	1	R96552	Hepatitis C virus type	1.38e+02
19	69	17.3	434	1	R80167	Mouse signal transduct	1.38e+02
20	69	17.3	434	1	R80220	GRB-7 adaptor protein	1.38e+02
21	69	17.3	435	1	R80161	GRB-7 central BLM doma	1.38e+02
22	69	17.3	530	1	W28591	Birch pollen ro factor	1.38e+02
23	69	17.3	534	1	R80164	Mouse signal transduct	1.38e+02
24	69	17.3	535	1	R96997	Human IRS-1	[illegible]
25	69	17.3	604	1	W10640	Membrane ty;	[illegible]
26	69	17.3	650	1	R71382	Drosophila	[illegible]
27	69	17.3	694	1	W31267	Drosophila	[illegible]
28	69	17.3	1728	1	R15144	Deleted in	[illegible]
29	68	17.1	84	1	R23417	Bovine para	[illegible]
30	68	17.1	275	1	W21788	Protein enc	[illegible]
31	68	17.1	294	1	R86456	Haloacetate	[illegible]
32	68	17.1	412	1	W71469	Death assoc	[illegible]
33	68	17.1	412	1	R74267	Human death	[illegible]
34	68	17.1	550	1	W69551	Brevibacter	[illegible]
35	68	17.1	550	1	W47498	B. lactofer	[illegible]
36	68	17.1	550	1	W25252	Brevibacter	[illegible]
37	68	17.1	550	1	W69584	B. lactofer	[illegible]
38	68	17.1	550	1	W68149	Brevibacter	[illegible]
39	68	17.1	711	1	W22169	ApxIIIC pr	[illegible]
40	68	17.1	880	1	P90855	Sequence c	[illegible]
41	67	16.8	84	1	R23455	Porcine pa	[illegible]
42	67	16.8	192	1	Y10093	HIV 1 vira	[illegible]
43	67	16.8	209	1	W53625	Fragment of	[illegible]
44	67	16.8	425	1	W47151	Human PMP-	[illegible]
45	67	16.8	452	1	W71852	N-acetylche	[illegible]

ALIGNMENTS

RESULT 1
ID P70138 standard; protein; 536 AA.
AC P70138;
DT 09-APR-1991 (first entry)
DE Sequence of a region of the S1 protein of infecti
DE bronchitis virus (IBV) strain D1466.
KW Poultry vaccine; antigen; immunogen.
OS Infectious bronchitis virus.
PN EP-221609-A.
PD 13-MAY-1987.
PF 28-OCT-1986; 201884.
PR 31-OCT-1985; NL-8502975.
PA (DUIN) DUPHAR INT RES BV.
PI Niesters HGM, Van der Zeijst BAM;.
DR WPI; 87-130709/19.
PT New gene coding for peptide - infectious bronchit
PT useful in prodn. of peptide - protein for vaccin.
PT against the viruses.
PS Claim 2; Fig 2; 21pp; Englist.
CC Proteins encoded by IBV strains M41, D1466, Beau-7
CC together with the genes that encode them, and the
CC a protein from IBV strain B1/20. In addition, a p
CC characterised in that it comprises at least one o
CC which the AA sequences of the IBV strains M41 and
CC different, esp. the region between AAs 55 and 70
CC 116 and 131 and/or between AAs 290 and 401.
SQ Sequence 536 AA;

Query Match [illegible] Score 86; DB 1; Len
Best Local Similarity [illegible]; Pred. No. 4.84e+
Matches 12; Conservative 1; Mismatches 10;

Db 44 AYLVINETDI SYNIVSFIVI IIKSGIVINESA-ISFVIKIFIA
[illegible]
Qy 5 GYLSRNSSRVCNYHGC[illegible]NSNQLSISMDSPVEW

RESULT 2
ID W81733 standard; Protein; 284 AA.
AC W81733;
DT 27 JAN 1999 (first entry)
DE M. tuberculosis immunogenic protein/peptide XP14 prot
KW Tuberculosis; immunogenic; soluble; antigen; prote
KW vaccine; pharmaceutical; infection; diagnosis.
OS Mycobacterium tuberculosis.
PN WO9816646-A2.
PD 23-APR-1998.

PF 07-OCT-1997; U18293.
PR 13-MAR-1997; US-818112.
PR 11-OCT-1996; US-730510.
PA (CORI-) CORIXA CORP.
PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ,
PI Reed SC, Skeiky YAW, Twardzik DR, Vedvick TS;
DR WPI; 98-261042/23.
DR N-PSDB; V64524.
PT Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA used
PT to develop products for the detection of M. tuberculosis infection
PT and for diagnosis, treatment and prevention of tuberculosis
PS Example 3c; Page 166; 230pp; English.
CC This sequence represents an immunogenic portion of a soluble
CC Mycobacterium tuberculosis (MT) antigen which can be used in a method
CC for inducing protective immunity against tuberculosis (TB). This sequence
CC can be formulated into vaccines and/or pharmaceutical compositions for
CC immunising against M. tuberculosis infection or may be used for the
CC diagnosis of tuberculosis.
SQ Sequence 284 AA;

Query Match 18.8%; Score 75; DB 1; Length 284;
Best Local Similarity 31.3%; Pred. No. 4.36e+01;
Matches 10; Conservative 12; Mismatches 8; Indels 2; Gaps 2;

Db 64 VGFLEPAPVLVDQRDDLGGLTVDWKVSWPRQR 95
:|:||:: : : ::|: |::| | | |:|
Qy 24 IGILEAGQIYGN-SNQLS-LSMDSPVEWRRRR 53

RESULT 3
ID W64366 standard; Protein; 284 AA.
AC W64366;
DT 09-NOV-1998 (first entry)
DE Mycobacterium tuberculosis polypeptide.
KW Tuberculosis; infection; diagnosis; antigen; XP14.
OS Mycobacterium tuberculosis strain Erdman.
PN WO9816645-A2.
PD 23-APR-1998.
PF 07-OCT-1997; U18214.
PR 13-MAR-1997; US-818111.
PR 11-OCT-1996; US-729622.
PA (CORI-) CORIXA CORP.
PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ,
PI Reed SC, Skeiky YAW, Twardzik DR, Vedvick TS;
DR WPI; 98-251292/22.
DR N-PSDB; V44415.
PT New isolated Mycobacterium tuberculosis polypeptides and DNA - used
PT to develop products for the detection of M. tuberculosis infection
PT and diagnosis of tuberculosis
PS Example 3; Page 175-176; 250pp; English.
CC This polypeptide is encoded by a DNA molecule (see V44415) isolated
CC from a Mycobacterium tuberculosis strain Erdman genomic DNA
CC expression library using sera from patients having extrapulmonary
CC tuberculosis. The invention relates to compositions and methods
CC for diagnosing tuberculosis. It provides polypeptides (see
CC W64291 W64379) comprising an antigenic portion of a soluble
CC M. tuberculosis antigen, or an immunogenic portion of a M.
CC tuberculosis antigen, as well as DNA sequences encoding such
CC polypeptides, recombinant expression vectors and transformed or
CC transfected host cells. Also claimed are methods and diagnostic
CC kits for detecting M. tuberculosis infection in a patient using
CC these polypeptides, antibodies or oligonucleotide probes and
CC primers.
SQ Sequence 284 AA;

Query Match 18.8%; Score 75; DB 1; Length 284;
Best Local Similarity 31.3%; Pred. No. 4.36e+01;
Matches 10; Conservative 12; Mismatches 8; Indels 2; Gaps 2;

Db 64 VGFLEPAPVLVDQRDDLGGLTVDWKVSWPRQR 95
:|:||:: : : ::|: |::| | | |:|
Qy 24 IGILEAGQIYGN-SNQLS-LSMDSPVEWRRRR 53

RESULT 4
ID W18664 standard; Protein; 386 AA.
AC W18664;
DT 24-JUL-1997 (first entry)
DE Fragmented human NF-H gene +1 frameshift mutant product.
KW Frameshift mutation product; GAGA motif; somatic mutation; diagnosis;
KW detection; antibody; probe; cancer; neoplasia; neurodegenerative;
KW Parkinson's, Alzheimer's disease, Pick's, Huntington's disease;
KW Down's syndrome; frontal lobe dementia; progressive supranuclear palsy;
KW PSP; amyotrophic lateral sclerosis, multiple sclerosis; MS;
KW cardiovascular; rheumatoid arthritis.
OS Homo sapiens.
FH Key Location/Qualifiers
FT misc_difference 1. .387
FT /note= "X corresponds to a stop codon in the
FT accompanying DNA file, T69796"
FT peptide 129. .135
FT /note= "antigenic peptide used fro antibody
FT production"
PN WO9712982-A2.
PD 10-APR-1997.
PF 02-OCT-1996; IB1106.
PR 02-OCT-1995; GB-020080.
PR 11-JAN-1996; US-009832.
PA (ROYA-) ROYAL NETHERLANDS ACAD ARTS & SCI.
PA (UYRO-) UNIV ROTTERDAM ERASMUS.
PA (UYUT-) UNIV STATE UTRECHT.
PI Burbach JPH, Grosveld FG, Van Leeuwen FW;
DR WPI; 97-226235/20.
DR N-PSDB; T69796.
PT Use of mutant genes having frameshift mutation(s) - for developing
PT prods. for the diagnosis, prevention and treatment of associated
PT diseases, e.g. cancer or neurodegenerative disease
PS Claim 22; Fig 9; 123pp; English.
CC W18663 and W18664 are +2 and +1 frameshift mutations, respectively,
CC of a sequence comprising fragments of the coding sequence of the
CC human neurofilament subunit NF-H gene corresponding to nucleotides
CC 1-1162 of the wild-type NF-H gene. This region contains GAGAG motifs.
CC Frameshift mutants of the tau, ubiquitin, apolipoprotein E,
CC microtubule-associated protein 2 (MAP-2), neurofilament subunit L, M
CC and H and amyloid A4 genes are claimed. All these genes share a common
CC GAGAN motif (N= A, G, C or T), which is the site of common GA
CC dinucleotide deletion(s) that cause neurodegenerative disorders.
CC Antigenic peptides used for the production of antibodies, and small
CC nucleic acid sequences derived from frameshift mutants are used in the
CC diagnosis, prevention and treatment of cancer and neurodegenerative
CC diseases, e.g. Parkinson's disease, Alzheimer's disease, Down's
CC syndrome, frontal lobe dementia (Pick's disease), progressive
CC supranuclear palsy (PSP), amyotrophic lateral sclerosis, Huntington's
CC disease, multiple sclerosis, and other degenerative diseases such as
CC cardiovascular disease and rheumatoid arthritis.
SQ Sequence 386 AA;

Query Match 18.8%; Score 75; DB 1; Length 386;
Best Local Similarity 41.2%; Pred. No. 4.36e+01;
Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Db 148 ELRVRLAQPLPWRRRRL 164
:| : : |: |||||:
Qy 48 QLSLSMDSPVEWRRRRI 54

RESULT 5
ID R21200 standard; Protein; 84 AA.
AC R21200;
DT 17-JUN-1992 (first entry)
DE Human parathyroid hormone analogue, [Leu3Ala6]hPTH.
KW PTH; agonist; antagonist; receptor binding.
OS Synthetic.
PN WO9200753-A.
PD 23-JAN-1992.
PF 15-JUL-1990; U04971.

PR 13-JUL-1990; US-553760.
PA (REGC) Univ of California.
PI Cohen FA, Nissenson RA, Strewler GJ;
DR WPI; 92-056643/07.
PT New parathyroid hormone analogues - useful in treating cancer,
PT osteoporosis, hypercalcaemia and hyper-parathyroid conditions
PS Claim 15; Page 64; 86pp; English.
CC Residues 35-84 may be absent. The C-terminal gp. may be -COOH,
CC -COO-+M (M+ = cation), or -(C=O)NH2. 140 specific peptides derived
CC from human PTH are given in the specification (R21150-256, R23226-
CC 3250 + R23522-529). Corresp. peptides created using residues 7-84
CC of bovine and porcine PTH are also claimed. All have mutations at
CC positions 3, and/or 6, and/or 9 which result in surface side chains
CC which are useful to modulate receptor binding and activity. They
CC are useful as agonists and antagonists in the treatment of condi-
CC tions or diseases involving PTH. The peptides are pref. prepd. by
CC solid phase synthesis. See also R21257 (human generic), R21258
CC (bovine generic) and R21259 (porcine generic).
SQ Sequence 84 AA;

Query Match 17.8%; Score 71; DB 1; Length 84;
Best Local Similarity 30.0%; Pred. No. 9.43e+01;
Matches 9; Conservative 12; Mismatches 7; Indels 2; Gaps 2;

Db 1 SVLEIALMHNLGKHLN-SMER-VEWLRKKL 28
::|| : : :::| ||: |||
Qy 25 GILEAGQIYGNSNQLSLSMDSPVEWKRRRI 54

RESULT 6
ID R21214 standard; Protein; 84 AA.
AC R21214;
DT 17-JUN-1992 (first entry)
DE Human parathyroid hormone analogue, [Leu3Ala6]hPTH.
KW PTH; agonist; antagonist; receptor binding.
OS Synthetic.
PN WO9200754-A.
PD 23-JAN-1992.
PF 15-JUL-1990; 004971.
PR 13-JUL-1990; US-553760.
PA (REGC) Univ of California.
PI Cohen FA, Nissenson RA, Strewler GJ;
DR WPI; 92-056643/07.
PT New parathyroid hormone analogues - useful in treating cancer,
PT osteoporosis, hypercalcaemia and hyper-parathyroid conditions.
PS Claim 15; Page 65; 86pp; English.
CC Residues 35-84 may be absent. The C-terminal gp. may be -COOH,
CC -COO-+M (M+ = cation), or -(C=O)NH2. 140 specific peptides derived
CC from human PTH are given in the specification (R21150-256, R23226-
CC 3250 + R23522-529). Corresp. peptides created using residues 7-84
CC of bovine and porcine PTH are also claimed. All have mutations at
CC positions 3, and/or 6, and/or 9 which result in surface side chains
CC which are useful to modulate receptor binding and activity. They
CC are useful as agonists and antagonists in the treatment of condi-
CC tions or diseases involving PTH. The peptides are pref. prepd. by
CC solid phase synthesis. See also R21257 (human generic), R21258
CC (bovine generic) and R21259 (porcine generic).
SQ Sequence 84 AA;

Query Match 17.8%; Score 71; DB 1; Length 84;
Best Local Similarity 30.0%; Pred. No. 9.43e+01;
Matches 9; Conservative 12; Mismatches 7; Indels 2; Gaps 2;

Db 1 SVLEIALMHNLGKHLN-SMER-VEWLRKKL 28
::|| : : :::| ||: |||
Qy 25 GILEAGQIYGNSNQLSLSMDSPVEWKRRRI 54

RESULT 7
ID P61418 standard; Protein; 319 AA.
AC P61418;
DT 28-OCT-1991 (first entry)
DE Serratia phospholipase protein.
KW Phospholipids; fatty acids; ;NOT21.
OS Serratia sp.
PN WO8606743-A.
PD 20-NOV-1986.
PF 09-MAY-1986; DK0051.
PR 10-MAY-1985; DK-002150.
PR 23-DEC-1985; DK-006060.
PR 09-MAY-1986; WO-DK00051.
PR 09-JAN-1987; DK-000047.
PA (BENA) BENZON A AS.
PA (MOLI/) MOLIN S.
PI Molin S, Givskov M, Kilse E;
DR WPI; 86-328851/48.
DR N-PSDB; N60877.
PT Producing bacterial enzyme esp. nuclease - using
PT which contains DNA from Serratia spp. encoding ex
PT enzyme.
PS Disclosure; Fig 9; 66pp; English.
CC The gene product may be expressed by an E.coli host
CC a plasmid encoding the phospholipase gene esp. und
CC Serratia regulatory sequences. The phospholipase g
CC be used to hydrolyse fatty acids from phospholipi
SQ Sequence 319 AA;

Query Match 17.8%; Score 71; DB 1; Leng
Best Local Similarity 27.3%; Pred. No. 9.43e+01;
Matches 9; Conservative 7; Mismatches 7;

Db 131 FQAG-IYSNDKQYVLAFACINAWR 15
::|| |:|: :| |: : ::
Qy 27 LEAGQIYGNSNQLSLSMDSPVEWR 50

RESULT 8
ID W20728 standard; protein; 31 AA.
AC W20728;
DT 15-JUL-1997 (first entry)
DE H. pylori cytoplasmic protein, 3-cel0515or14.
KW Vaccine, prevention, treatment; infection; identif
KW binding compound; bacterium; life cycle; activator
KW inhibitor; duodenal ulcer disease; chronic gastrit
KW cytoplasmic; mRNA; translation; ribosome; biogenes
OS Helicobacter pylori.
PN WO9640893-A1.
PD 19-DEC-1996.
PF 06-JUN-1996; 009122.
PR 07-JUN-1995; US-487043.
PR 01-APR-1996; US 630405.
PA (ASTR) ASTRA AB.
PI Berglindh OT, Smith K, Mellgard BL;
DR WPI; 97-052306/05.
DR N-PSDB; I67981.
PT Helicobacter pylori nucleic acid sequences and for
PT polypeptide(s) useful for vaccines to treat or
PT infection, and to detect Helicobacter
PS Claim 61; Pages 1143-1144, 1436pp; English.
CC The present sequence is a helicobacter pylori cyto
CC involved in mRNA translation and ribosome biogenes
CC The protein may be used in a vaccine to prevent or
CC H. pylori infection or to identify H. pylori polyp
CC compounds, useful as potential H. pylori life cycl
CC inhibitors. The genomic sequence of H. pylori (ATC
CC determined from overlapping contigs generated by m
CC shearing the bacterial DNA. The sequences were ana
CC at least 180 nucleotides, and the predicted coding
CC by computer evaluation. To identify likely H. pylor
CC vaccine development, the amino acid sequences pred.
CC various ORF were analysed for significant homology
CC or exported membrane proteins. Having identified a
CC the sequences of interest, particular regions can
CC H. pylori by PCR amplification for recombinant poly
CC production, e.g. in E. coli hosts.
SQ Sequence 319 AA;

```
  Query Match             17.8%;  Score 71;  DB 1;  Length 319;
  Best Local Similarity   31.9%;  Pred. No. 9.43e+01;
  Matches    15;  Conservative    12;  Mismatches   18;  Indels    2,  Gaps    2,

Db     114 LFIAPCINLHASLLPKYRGASPIHEMILNDDRIYGISIMLMDLELDS 160
           : |    :: ::| : :|:| |:    ||:  :||| |  |  | :||
Qy       1 MVIEGYLSRNSSRVCNYHGCSSAIG-ILEAGQIYGNSNQL-SLSMDS 45


RESULT    9
ID   W90188 standard; Protein; 449 AA.
AC   W90188;
DT   10-MAY-1999  (first entry)
DE   Arabidopsis Golgi glycosyltransferase qtl6.
KW   Glycosyltransferase; Golgi; qtl6; transgenic plant, xylose, lateral,
KW   plant cell wall; paper; pulp; food.
OS   Arabidopsis thaliana.
FH   Key             Location/Qualifiers
FT   Domain          33. .51
FT                   /note= "transmembrane domain"
FT   Domain          130. .330
FT                   /note= "catalytic domain"
PN   WO9901558-A1.
PD   14-JAN-1999
PF   30-JUN-1998; 001911.
PR   30-JUN-1997; GB-013852.
PA   (UYCA-) UNIV CAMBRIDGE TECH SERVICES LTD.
PI   Dupree P, Moqelsvang S;
DR   WPI; 99-106058/09.
DR   N-PSDB; V74278.
PT   New isolated plant glycosyltransferase nucleic acids  used to
PT   transform plants to alter the properties or composition of cell
PT   walls, e.g. for modifying foodstuffs or wood pulp products
PS   Claim 8; Fig 1A-C; 66pp; English.
CC   This is the predicted amino acid sequence of a novel Golgi
CC   glycosyltransferase, termed qtl6 (qma twelve-like 6), of
CC   Arabidopsis thaliana.  Full-length qtl6 cDNA (see V74278) was
CC   isolated following an expressed sequence tag (EST) database
CC   screening and 5'RACE.  The predicted protein has all the
CC   characteristics of a Golgi glycosyltransferase, i.e. a short
CC   cytosolic N-terminus, single transmembrane domain, and a larger
CC   lumenal C-terminal domain with conserved cysteines.  The plant
CC   glycosyltransferases of the invention (see W99188-94) can catalyse
CC   one or more steps in the biosynthesis of complex non-cellulosic
CC   plant cell wall polysaccharides.  They can also catalyse the
CC   glycosylation of plant cell wall protein and/or the transfer of
CC   saccharide moieties from nucleotide sugar precursors to
CC   oligosaccharide precursors.  Nucleic acids encoding the
CC   glycosyltransferases can be used to transform plants to alter the
CC   properties or composition of the cell wall.  The plants produced
CC   can be used in the preparation of e.g. food or wood pulp products
CC   such as paper or cardboard.
SQ   Sequence   449 AA;

  Query Match             17.8%;  Score 71;  DB 1;  Length 449;
  Best Local Similarity   30.8%;  Pred. No. 9.43e+01;
  Matches    12;  Conservative    13;  Mismatches    9;  Indels    5,  Gaps    4;

Db      71 CSSGIDMSQDPIDPVYYDDPDLPYTIEKPVKNWDEKRRR 109
           |||:| :  : :  :| :  :|: ::: || :|  :|||
Qy      20 CSSAIGI-LE-AGQIYGNSNQLSLSMDSPV-EW--RRRR 53


RESULT   10
ID   W89248 standard; Protein; 488 AA.
AC   W89248;
DT   10-MAR-1999  (first entry)
DE   Human SAD.
KW   PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;
KW   type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;
KW   neurodegenerative disease; neuronal survival; Alzheimer's disease;
KW   Parkinson's disease; Huntington's disease.
OS   Homo sapiens.
PN   WO9849317-A2.
PD   05-NOV-1998.
DT   27-APR-1998; 008439.
PR   23-OCT-1997; US-063595.
PR   28-APR-1997; US-044428.
PR   29-MAY-1997; US-047222.
PR   11-JUN-1997; US-049477.
PR   11-JUN-1997; US-049756.
PR   18-JUN-1997; US-049914.
PA   (SUGE-) SUGEN INC.
PI   App H, Clary D, Courtneidge SA, Hui TH, Jallal B,
PI   Markby D, Onrust S, Peles E, Plowman GD;
DR   WPI; 99-009434/01.
DR   N-PSDB; V81743.
PT   New nucleic acid encoding specific protein tyrosine phosphatases -
PT   useful for identifying specific modulators for treatment and
PT   prevention of cancer and neurodegenerative disease
PS   Claim 2; Page 154-155; 193pp; English.
CC   The present invention describes isolated, enriched or purified nucleic
CC   acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The
CC   present sequence represents human SAD. The above proteins, other than
CC   ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify
CC   substances that modulate their activity (i.e. agonists and antagonists,
CC   including NBP) in vivo or in vitro. These substances are used to treat
CC   or prevent diseases associated with abnormal signal transduction
CC   pathways that involve the proteins, particularly cancer (e.g. leukaemia
CC   and lymphoma), while modulators of ALK-7 (which is a type I receptor
CC   serine/threonine kinase) are used to promote neuronal survival,
CC   particularly for treating Alzheimer's, Parkinson's or Huntington's
CC   diseases. Nucleic acid fragments of the polynucleotides encoding the
CC   proteins can be used as probes to identify and clone related sequences;
CC   to detect protein-encoded RNA; to generate transgenic animals and in
CC   gene therapy (optionally after mutation). Ab are used to determine the
CC   proteins.
SQ   Sequence   488 AA;

  Query Match             17.8%;  Score 71;  DB 1;  Length 488;
  Best Local Similarity   31.7%;  Pred. No. 9.43e+01;
  Matches    13;  Conservative    13;  Mismatches   11;  Indels    4;  Gaps    4;

Db     153 LGGYSLSVRAQAKVCHYRVSMAADGSLYLQKGRLFFCLEEL 193
           : || || |  ::||:|:   :| | : |: |:::   ::|
Qy       8 IEGY-LS-RNSSRVCNYHGCSSAIG-I-LEAGQIYGNSNQL 39


RESULT   11
ID   R74171 standard; Protein; 3038 AA.
AC   R74171;
DT   18-JAN-1996  (first entry)
DE   Aspergillus terreus triol polyketide synthase.
KW   Triol polyketide synthase; TPKS; HMG-CoA reductase inhibitor;
KW   hypercholesterolaemia; LDL- cholesterol.
OS   Aspergillus terreus.
FH   Key             Location/Qualifiers
FT   region          181
FT                   /label= keto-acyl synthase motif
FT   region          654. .658
FT                   /label= acetyl/malonyl transferase motif
FT   region          985. .994
FT                   /label= dehydratase motif
FT   region          1446. .1450
FT                   /label= methyl transferase motif
FT   region          1932. .1937
FT                   /label= enoyl reductase motif
FT   region          2164. .2169
FT                   /label= keto reductase motif
FT   region          2498
FT                   /label= acyl carrier protein motif
FT   misc_difference 282. .288
FT                   /label= misc feature
FT   misc_difference 1450. .1460
FT                   /label= misc feature
```

```
FT   misc_difference 1603. .1612
FT                   /label= misc feature
FT   misc_difference 2521. .2535
FT                   /label= misc feature
PN   WO9512661-A.
PD   11-MAY-1995.
PF   28-OCT-1994; 94WO-US12423.
PR   02-NOV-1993; 93US-148132.
PA   (MERI ) MERCK & CO INC.
PI   Condon MJ, Davis CR, Hendrickson TF, Meada PC, Pambrosek J;
PI   Reeves CD, Vinci VA;
DR   WPI; 95-193816/25.
DR   N-PSDB; Q92328.
PI   Novel DNA encoding triol poly-ketide synthase - used to isolate and
PI   identify homologues of triol poly-ketide synthase, and in the treatment
PI   of hyper-cholesteroiaemia
PS   Claim 12; Figure 2; 107pp; English.
CC   The full-length TPKS-encoding DNA in plasmid pLOA was
CC   designated pTPKS100. Splicing of the introns from the DNA
CC   sequence and translation of the 9114 nt ORF results in a
CC   protein of 3038 AAs (R74171) with a mol. wt. of 269,090
CC   daltons. Inspection of the TPKS AA sequence for active
CC   site residues and motifs known to be associated with
CC   polyketide synthases and fatty acid synthase (FAS) activities
CC   resulted in the identification of candidates for expected
CC   sites (see FT). Except for the presence of a methyl transferase,
CC   not present in FAS, the succession of activities on the
CC   TPKS protein is the same as that observed for the rat FAS
CC   protein.
SQ   Sequence   3038 AA;

  Query Match             17.8%;  Score 71;  DB 1;  Length 3038;
  Best Local Similarity  29.5%;  Pred. No. 9.4e+01;
  Matches   13;  Conservative   11;  Mismatches   10;  Indels   2;  Gaps   2;

Db    444 IIEEYMEPEQNQLRVSNNEDCPPMTGVLSLPLVLSAKSQRSLKI 487
          :|| |: : :|  || |   |:: :|:|   : :  :| || :
Qy      2 VIEGYL-S-RNSSRVCNYHSCSSAISILFAGQIYGNSNQLSLSM 43


RESULT   12
ID   R23358 standard; Protein; 84 AA.
AC   R23358;
DT   22-JUN-1992  (first entry)
DE   Bovine parathyroid hormone analogue, [Ser1Leu3Ala6]bPTH.
KW   PTH; agonist; antagonist; receptor binding.
OS   Synthetic.
PN   WO9200753-A.
PD   23-JAN-1992.
PF   15-JUL-1990; 90WO-US04071.
PR   13-JUL-1990; 90US-553760.
PA   (REGC ) Univ of California.
PI   Cohen FA, Nissenson RA, Strewler GJ;
DR   WPI; 92-056643/07.
PT   New parathyroid hormone analogues - useful in treating cancer,
PT   osteoporosis, hypercalcaemia and hyper-parathyroid condi ions.
PS   Claim 15; Page 65; 86pp; English.
CC   Residues 35-84 may be absent. The C terminal gp. may be -COOH,
CC   -COO-+M (M+ = cation), or -(C=O)NH2. 142 specific peptides derived
CC   from bovine PTH are given in the specification (R23251-382, and
CC   R23540-549). Corresp. peptides created using residues 7 84
CC   of human and porcine PTH are also claimed. All have mutations at
CC   positions 3, and/or 6, and/or 9 which result in surface side chains
CC   which are useful to modulate receptor binding and activity. They
CC   are useful as agonists and antagonists in the treatment of condi-
CC   tions or diseases involving PTH. The peptides are pret. prepd. by
CC   solid phase synthesis. See also R21257 (human generic), R21258
CC   (bovine generic) and R21259 (porcine generic).
SQ   Sequence   84 AA;

  Query Match             17.6%;  Score 70;  DB 1;  Length 84;
  Best Local Similarity  33.3%;  Pred. No. 1.14e+02;
  Matches   10;  Conservative   11;  Mismatches   7;  Indels   2;  Gaps   2;

Db      1 SVLEIAFMHNLGKHLS SMER VEWLRKKL 28
          :::| : :   :::|| ||:  | |  :::
Qy     25 GILEAGQIYGNSNQLSLSMERPVEWRFFRI 54


RESULT   13
ID   R23344 standard; Protein; 84 AA.
AC   R23344;
DT   22-JUN-1992  (first entry)
DE   Bovine parathyroid hormone analogue, [Ser1Leu3Ala
KW   PTH; agonist; antagonist; receptor binding.
OS   Synthetic.
PN   WO9200753-A.
PD   23-JAN-1992.
PF   15-JUL-1990; 90WO-US04071.
PR   13-JUL-1990; 90US-553760.
PA   (REGC ) Univ of California.
PI   Cohen FA, Nissenson RA, Strewler GJ;
DR   WPI; 92-056643/07.
PT   New parathyroid hormone analogues - useful in treat
PT   osteoporosis, hypercalcaemia and hyper-parathyroid
PS   Claim 15; Page 64; 86pp; English.
CC   Residues 35-84 may be absent. The C-terminal gp. m
CC   -COO-+M (M+ = cation), or -(C=O)NH2. 142 specific
CC   from bovine PTH are given in the specification (R2
CC   R23540-549). Corresp. peptides created using resi
CC   of human and porcine PTH are also claimed. All have
CC   positions 3, and/or 6, and/or 9 which result in su
CC   which are useful to modulate receptor binding and
CC   are useful as agonists and antagonists in the treat
CC   tions or diseases involving PTH. The peptides are
CC   solid phase synthesis. See also R21257 (human gene
CC   (bovine generic) and R21259 (porcine generic).
SQ   Sequence   84 AA;

  Query Match             17.6%;  Score 70;  DB 1;  Length
  Best Local Similarity  33.3%;  Pred. No. 1.14e+02;
  Matches   10;  Conservative   11;  Mismatches   7;

Db      1 SVLEIAFMHNLGKHLS SMER VEWLRKKL 28
          :::| : :   :::|| ||:  | |  :::
Qy     25 GILEAGQIYGNSNQLSLSMERPVEWRFFRI 54


RESULT   14
ID   W44269 standard; Protein; ? AA.
AC   W44269;
DT   19 MAY-1998  (first entry)
DE   Hybrid DNA protein comprising a family 45 cellulase
KW   Cellulase core region; isolated microorganism; h
KW   hybrid DNA.
OS   Chimeric - Humicola nigrescens.
OS   Chimeric - Humicola insolens.
PN   WO9743409-A2.
PD   20-NOV-1997.
PF   12-MAY-1997; 97WO-DK00216.
PR   10-MAY-1996; 96DK-000562.
PA   (NOVO ) NOVO NORDISK AS.
PI   Dalboge H, Diderichsen B, Kauppinen S, Sandal I;
DR   WPI; 98-008878/01.
DR   N-PSDB; W15072.
PI   Isolating novel DNA sequences from microorganisms
PI   for culturing the microorganism
PS   Example 3; Page 52-54; 72pp; English.
CC   The present sequence represents a protein encoded f
CC   from an example of the present invention. The prese
CC   describes a novel method for providing a novel DNA
CC   polypeptide from a microorganism with an activity
CC   method comprises: (i) PCR amplification of the DNA
CC   homology to (a) known gene(s) encoding a polypeptid
CC   interest; (ii) linking the obtained PCR product of
CC   sequence and a 3' structural gene sequence; (iii)
```

```
CC   resulting hybrid DNA sequence; (iv) screening for hybrid DNA sequences
CC   encoding a polypeptide with the activity of interest or a related
CC   activity; and (v) isolating the hybrid DNA sequence identified in step
CC   (iv). This method provides for identification and isolation of sequences
CC   from microorganisms without having to cultivate and isolate the
CC   microorganism.
SQ   Sequence   306 AA;

  Query Match           17.6%;  Score 70;  DB 1;  Length 306;
  Best Local Similarity  42.9%;  Pred. No. 1.14e+02;
  Matches   12;  Conservative    6;  Mismatches   9;  Indels   1;  Gaps   1;

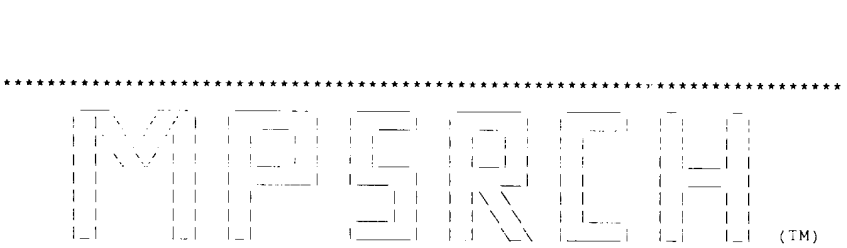
Db     144 LNMPGGGVGIFDGCSPQVGGL-AGQRYG 170
           |:  :: |  : |||: :| | ||| ||
Qy       7 LSRNSSRVGNYHGCSSAIGILEAGQIYG 34


RESULT   15
ID   R23541 standard; Protein; 84 AA.
AC   R23541;
DT   22-JUN-1992  (first entry)
DE   Bovine parathyroid hormone analogue, [Tyr9]bPTH.
KW   PTH; agonist; antagonist; receptor binding.
OS   Synthetic.
PN   WO9200753-A.
PD   23-JAN-1992.
PF   15-JUL-1990; 004971.
PR   13-JUL-1990; US-553760.
PA   (REGC ) Univ of California.
PI   Cohen FA, Nissenson RA, Strewler GJ;
DR   WPI; 92-056643/07.
PT   New parathyroid hormone analogues - useful in treating cancer,
PT   osteoporosis, hypercalcaemia and hyper-parathyroid conditions.
PS   Claim 15; Page 66; 86pp; English.
CC   Residues 35-84 may be absent. The C-terminal gp. may be -COOH,
CC   -COO-.M (M+ = cation), or -(C=O)NH2.  142 specific peptides derived
CC   from bovine PTH are given in the specification (R23251-382, and
CC   R23540-549).  Corresp. peptides created using residues 7-84
CC   of human and porcine PTH are also claimed. All have mutations at
CC   positions 3, and/or 6, and/or 9 which result in surface side chains
CC   which are useful to modulate receptor binding and activity.  They
CC   are useful as agonists and antagonists in the treatment of condi-
CC   tions or diseases involving PTH. The peptides are pref. prepd. by
CC   solid phase synthesis. See also R21257 (human generic), R21258
CC   (bovine generic) and R21259 (porcine generic).
SQ   Sequence   84 AA;

  Query Match           17.3%;  Score 69;  DB 1;  Length 84;
  Best Local Similarity  39.1%;  Pred. No. 1.38e+02;
  Matches    9;  Conservative    8;  Mismatches   4;  Indels   2;  Gaps   2;

Db       8 MYNLGKRLS-SMER-VEWLRKKL 28
           :|  :::|| ||:  ||| |:::
Qy      42 IYGNSNQLSLSMDSPVEWRRRI 54


Search completed: Fri Aug 25 18:07:04 2000
Job time : 8 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 25 18:09:28 2000; MasPar time 2.62 Seconds
416.235 Million cell updates/sec

Tabular output not generated.

Title: >US-09-257-585-8
Description: (1-54) from US09257585.pep
Perfect Score: 398
Sequence: 1 MVIEGYLSRNSSRVCNYHGC NSNQLSLSMDSIVEWRRRPI 54

Scoring table: PAM 150
Gap 11

Searched: 152433 seqs, 15329240 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:6_COMB 4:PCI_COMB 5:backfiles1

Statistics: Mean 24.818; Variance 91.396; scale 0.272

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	75	18.8	188	2	US-08-726-	Sequence 57, Applicati	2.16e+01
2	71	17.8	203	2	US-08-640-	Sequence 18, Applicati	4.77e+01
3	71	17.8	3038	2	US-08-647	Sequence 2, Applicatio	4.77e+01
4	71	17.8	3038	1	US-08-450-	Sequence 2, Applicatio	4.77e+01
5	70	17.6	360	2	US-08-483-	Sequence 17, Applicati	5.80e+01
6	69	17.3	334	1	US-08-297	Sequence 9, Applicatio	7.04e+01
7	69	17.3	344	4	US-08-472-	Sequence 9, Applicatio	7.04e+01
8	69	17.3	345	4	US-08-866-	Sequence 1, Applicatio	7.04e+01
9	69	17.3	534	4	US-08-866-	Sequence 5, Applicatio	7.04e+01
10	69	17.3	535	1	US-07-906-	Sequence 10, Applicati	7.04e+01
11	69	17.3	535	1	US-08-167-	Sequence 10, Applicati	7.04e+01
12	69	17.3	535	2	US-08-549-	Sequence 10, Applicati	7.04e+01
13	69	17.3	535	1	US-08-208-	Sequence 10, Applicati	7.04e+01
14	69	17.3	650	1	US-08-121-	Sequence 60, Applicati	7.04e+01
15	69	17.3	650	4	US-08-833-	Sequence 60, Applicati	7.04e+01
16	69	17.3	650	1	US-08-835-	Sequence 60, Applicati	7.04e+01
17	69	17.3	650	4	PCT-US94-1	Sequence 60, Applicati	7.04e+01
18	69	17.3	650	2	US-09-060	Sequence 60, Applicati	7.04e+01
19	68	17.1	412	1	US-08-208-	Sequence 12, Applicati	8.54e+01
20	68	17.1	550	1	US-08-674-	Sequence 29, Applicati	8.54e+01
21	68	17.1	711	4	US-08-772-	Sequence 12, Applicati	8.54e+01
22	67	16.8	322	1	US-08-014	Sequence 2, Applicatio	1.03e+02
23	67	16.8	322	1	US-08-486-	Sequence 3, Applicatio	1.03e+02
24	67	16.8	322	2	US-08-486-	Sequence 3, Applicati	[illegible]
25	67	16.8	422	1	US-08-470-	Sequence 1, Applicati	[illegible]
26	67	16.8	591	4	US-08-965-	Sequence 2, Applicati	[illegible]
27	67	16.8	824	1	US-08-291-	Sequence 2, Applicati	[illegible]
28	67	16.8	828	2	US-08-485-	Sequence 2, Applicati	[illegible]
29	66	16.6	84	2	US-08-475-	Sequence 15, Applicati	[illegible]
30	66	16.6	86	1	US-07-956-	Sequence 24, Applicati	[illegible]
31	66	16.6	86	1	US-08-476-	Sequence 10, Applicati	[illegible]
32	66	16.6	86	1	US-08-485-	Sequence 15, Applicati	[illegible]
33	66	16.6	86	1	US-07-956-	Sequence 14, Applicati	[illegible]
34	66	16.6	86	1	US-08-485-	Sequence 24, Applicati	[illegible]
35	66	16.6	86	1	US-08-476-	Sequence 23, Applicati	[illegible]
36	66	16.6	86	2	US-08-475-	Sequence 23, Applicati	[illegible]
37	66	16.6	95	1	US-08-476-	Sequence 23, Applicati	[illegible]
38	66	16.6	95	1	US-07-956-	Sequence 23, Applicati	[illegible]
39	66	16.6	95	2	US-08-475-	Sequence 23, Applicati	[illegible]
40	66	16.6	95	1	US-08-485-	Sequence 23, Applicati	[illegible]
41	66	16.6	266	5	5269441-1	atent No. 5	[illegible]
42	66	16.6	344	4	US-08-691-	Sequence 1, Applicati	[illegible]
43	66	16.6	707	4	US-08-772-	Sequence 4, Applicati	[illegible]
44	66	16.6	2163	1	US-08-746-	Sequence 5, Applicati	[illegible]
45	65	16.3	1548	2	US-08-460-	Sequence 7, Applicati	[illegible]

ALIGNMENTS

RESULT 1
ID US-08-726-306A-57 STANDARD; PRT; 188 AA
XX
AC xxxxxx
XX
DT
XX
DE Sequence 57, Application US/08726306A
XX
CC Sequence 57, Application US/08726306A
CC Patent No. 5958684
CC GENERAL INFORMATION:
CC APPLICANT: van Leeuwen, Frederik Willem
CC APPLICANT: Burbach, Johannes Peter Henri
CC APPLICANT: Grosveld, Franklin G.
CC TITLE OF INVENTION: DIAGNOSTIC METHOD AND MEAN
CC NUMBER OF SEQUENCES: 189
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Banner & Witcoff, Ltd.
CC STREET: 1 Financial Center
CC CITY: Boston
CC STATE: MA
CC COUNTRY: US
CC ZIP: 02111
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb st
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WordPerfect 6.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/726,306A
CC FILING DATE: 02-Oct-1996
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: GB 9520080.4
CC FILING DATE: 02-Oct-1995
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/009,832
CC FILING DATE: 01-Jan-1996
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Williams, Ph.D., Kathleen M.
CC REGISTRATION NUMBER: 34,380
CC REFERENCE/DOCKET NUMBER: 96,048-A (3255/007
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 345-9100
CC TELEFAX: (617) 345-9111
CC INFORMATION FOR SEQ ID NO: 57:
CC SEQUENCE CHARACTERISTICS:

```
CC         LENGTH:  188 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   188 AA; 20763 MW; 143883 CN;

  Query Match            18.8%;  Score 75;  DB 2;  Length 188;
  Best Local Similarity  41.2%;  Pred. No. 2.16e+01;
  Matches    7;  Conservative    7;  Mismatches   5;  Indels   3;  Gaps   2;

Db    148 ELRVRLAQPLPWRREEL 164
          :| : :  |: ||||:
Qy     38 QLSLSMDSPVFWRRRRI 54


RESULT    2
ID   US-08-640-344-18      STANDARD;      PRT;   203 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 18, Application US/08640344
XX
CC    Sequence 18, Application US/08640344
CC    Patent No. 5824315
CC     GENERAL INFORMATION:
CC       APPLICANT:  NAG, BISHWAJIT
CC       APPLICANT:  MUKKU, PRAPHA
CC       APPLICANT:  DESHPANDE, SHRIKANT
CC       TITLE OF INVENTION:  IMPROVING BINDING AFFINITY OF ANTIGENIC
CC       TITLE OF INVENTION:  PEPTIDES FOR MHC MOLECULES
CC       NUMBER OF SEQUENCES:  18
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  TOWNSEND & TOWNSEND & CREW LLP
CC         STREET:  TWO EMBARCADERO CENTER, 8TH FLOOR
CC         CITY:  SAN FRANCISCO
CC         STATE:  CALIFORNIA
CC         COUNTRY:  U.S.A.
CC         ZIP:  94111-3834
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/640,344
CC         FILING DATE:  30-APR-1996
CC         CLASSIFICATION:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  STORELLA ESQ., JOHN R.
CC         REGISTRATION NUMBER:  32,944
CC         REFERENCE/DOCKET NUMBER:  14058.004800
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (415) 576-0200
CC         TELEFAX:  (415) 576-0300
CC     INFORMATION FOR SEQ ID NO:  18:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  203 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  peptide
SQ   SEQUENCE   203 AA; 23397 MW; 246506 CN;

  Query Match            17.8%;  Score 71;  DB 2;  Length 203;
  Best Local Similarity  37.5%;  Pred. No. 4.77e+01;
  Matches    9;  Conservative    6;  Mismatches   9;  Indels   0;  Gaps   0;

Db    167 SGEVYTCQVELSVTSPLTVEWRAR 190
          :|:::|    :||::    |||| |
Qy     29 AGQIYGNSNQLSLSMDSPVFWRRR 52


RESULT    3
ID   US-08-637-640-2      STANDARD;      PRT;   3038 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application US/08637640
XX
CC    Sequence 2, Application US/08637640
CC    Patent No. 5849541
CC     GENERAL INFORMATION:
CC       APPLICANT:  VINCI, VICTOR A.
CC       APPLICANT:  CONDER, MICHAEL J.
CC       APPLICANT:  MCADA, PHYLLIS C.
CC       APPLICANT:  REEVES, CHRISTOPHER D.
CC       APPLICANT:  DAVIS, CHARLES R.
CC       APPLICANT:  HENDRICKSON, LEE E.
CC       APPLICANT:  KAMBOSEK, JOHN
CC       TITLE OF INVENTION:  DNA ENCODING TRIOL POLYKETIDE SYNTHASE
CC       NUMBER OF SEQUENCES:  3
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  CHRISTINE E. CARTY
CC         STREET:  P.O. BOX 2000; 126 E. LINCOLN AVENUE
CC         CITY:  RAHWAY
CC         STATE:  NJ
CC         COUNTRY:  USA
CC         ZIP:  07065
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/637,640
CC         FILING DATE:  23-AUG-1996
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  08/148,132
CC         FILING DATE:  01-NOV-1993
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  CARTY, CHRISTINE E.
CC         REGISTRATION NUMBER:  36,099
CC         REFERENCE/DOCKET NUMBER:  19076
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  908-594-6734
CC         TELEFAX:  908-594-4720
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  3038 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
CC       HYPOTHETICAL:  YES
CC       ORIGINAL SOURCE:
CC         ORGANISM:  TPKS Protein
SQ   SEQUENCE   3038 AA; 335002 MW; 46682910 CN;

  Query Match            17.8%;  Score 71;  DB 2;  Length 3038;
  Best Local Similarity  29.5%;  Pred. No. 4.77e+01;
  Matches   13;  Conservative   11;  Mismatches  18;  Indels   2;  Gaps   2;

Db    444 IIEEYMPEQNQLRVSNNEDCPPMTGVLSLPLVLSAKSQPSLKI 487
          :|| |: : :|  || |   |::  |:|    : :  :| || :
Qy      2 VIEGYL G RNGSRVCNTHGCGSAIGTLEAGQIYGNSNQLSLSM 43


RESULT    4
ID   US-08-450-332-2      STANDARD;      PRT;   3038 AA.
XX
AC   xxxxxx
```

```
XX
DT
XX
DE   Sequence 2, Application US/08450332
XX
CC    Sequence 2, Application US/08450332
CC    Patent No. 5744450
CC     GENERAL INFORMATION:
CC       APPLICANT:  VINCI, VICTOR A.
CC       APPLICANT:  CONDER, MICHAEL J.
CC       APPLICANT:  MCADA, PHYLLIS C.
CC       APPLICANT:  REEVES, CHRISTOPHER D.
CC       APPLICANT:  DAVIS, CHARLES R
CC       APPLICANT:  HENDRICKSON, LEE E.
CC       APPLICANT:  RAMBOSEK, JOHN
CC       TITLE OF INVENTION:  DNA ENCODING TRIOL POLYKETIDE SYNTHASE
CC       NUMBER OF SEQUENCES:  3
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  CHRISTINE F. CARTY
CC         STREET:  125 E. LINCOLN AVENUE, P.O. Box 2000
CC         CITY:  RAHWAY
CC         STATE:  NJ
CC         COUNTRY:  USA
CC         ZIP:  07065-0907
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  Patentin Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/450,332
CC         FILING DATE:  25-MAY-1995
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  CARTY, CHRISTINE E.
CC         REGISTRATION NUMBER:  36,099
CC         REFERENCE/DOCKET NUMBER:  190760A
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  908-594-6734
CC         TELEFAX:  908-594-4720
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  3038 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
CC       HYPOTHETICAL:  YES
CC       ORIGINAL SOURCE:
CC         ORGANISM:  IPKS Protein
SQ   SEQUENCE   3038 AA; 335002 MW; 46682910 CN;

  Query Match             17.8%;  Score 71;  DB 1;  Length 3038;
  Best Local Similarity   29.5%;  Pred. No. 4.77e+01;
  Matches   14;  Conservative   11;  Mismatches  18;  Indels   2,  Gaps   2;

Db    444 IIEEYMEPEQNQLRVSNNPIGPPMIGVISIPIVISAKSQRSIKI 487
          :|| |: : :|  || |   |::  |:|    : :  :| || :
Qy      2 VIEGYL-S-RNSSRVCNYHGCSSAIGILEAGQIYGNSNQLSLSM 43


RESULT    5
ID   US-08-483-101-17     STANDARD;     PRT;   360 AA.
XX
AC   XXXXXX
XX
DT
XX
DE   Sequence 17, Application US/08483101
XX
CC    Sequence 17, Application US/08483101
CC    Patent No. 5942715
CC     GENERAL INFORMATION:
CC       APPLICANT:  Scott, Jane E.
CC       APPLICANT:  Froehlich, Barbara
CC       APPLICANT:  Carol, Judy
CC       TITLE OF INVENTION:  PSA Proteins and Coding
CC       NUMBER OF SEQUENCES:  17
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Greenlee and Winner, P.C.
CC         STREET:  5370 Manhattan Circle, Suite 201
CC         CITY:  Boulder
CC         STATE:  Colorado
CC         COUNTRY:  US
CC         ZIP:  80303
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  Patentin Release #1.0, Version #
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/483,101
CC         FILING DATE:  07-JUN-1995
CC         CLASSIFICATION:  424
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Ferber, Donna M.
CC         REGISTRATION NUMBER:  33878
CC         REFERENCE/DOCKET NUMBER:  6-95
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (303) 499-8080
CC         TELEFAX:  (303) 499-8089
CC     INFORMATION FOR SEQ ID NO:  17:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  360 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:
CC         TOPOLOGY:  unknown
CC       MOLECULE TYPE:  protein
CC       HYPOTHETICAL:  NO
SQ   SEQUENCE   360 AA; 39775 MW; 591439 CN;

  Query Match             17.6%;  Score 70;  DB 2;  Length
  Best Local Similarity   43.5%;  Pred. No. 5.80e+01;
  Matches   10;  Conservative    5;  Mismatches   8;

Db    229 YIGRNSVDMCFYDGYSINSSSLE 251
          |::|||  :| | | |:  : ::
Qy      6 YLSRNSSKVCNYHGCSSAIGILE 28


RESULT    6
ID   US-08-207-575A-9     STANDARD;     PRT;   434 AA.
XX
AC   XXXXXX
XX
DT
XX
DE   Sequence 9, Application US/08207575A
XX
CC    Sequence 9, Application US/08207575A
CC    Patent No. 6037134
CC     GENERAL INFORMATION:
CC       APPLICANT:  Margolis, Benjamin L.
CC       TITLE OF INVENTION:  METHODS AND COMPOSITIONS
CC       TITLE OF INVENTION:  OF BREAST CANCER
CC       NUMBER OF SEQUENCES:  21
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  FENNIE & KIMMONS LLP
CC         STREET:  1155 Avenue of the Americas
CC         CITY:  New York
CC         STATE:  New York
CC         COUNTRY:  U.S.A.
CC         ZIP:  10036-2711
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
```

```
CC          OPERATING SYSTEM:  PC-DOS/MS-DOS
CC          SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC        CURRENT APPLICATION DATA:
CC          APPLICATION NUMBER:  US/08/207,575A
CC          FILING DATE:  07-MAR-1994
CC          CLASSIFICATION:  435
CC        ATTORNEY/AGENT INFORMATION:
CC          NAME:  Coruzzi, Laura A.
CC          REGISTRATION NUMBER:  30,742
CC          REFERENCE/DOCKET NUMBER:  7683-053
CC        TELECOMMUNICATION INFORMATION:
CC          TELEPHONE:  (212) 790-9090
CC          TELEFAX:  (212) 869-9741/8864
CC          TELEX:  66141 PENNIE
CC      INFORMATION FOR SEQ ID NO:  9:
CC        SEQUENCE CHARACTERISTICS:
CC          LENGTH:  334 amino acids
CC          TYPE:  amino acid
CC          STRANDEDNESS:  single
CC          TOPOLOGY:  unknown
CC        MOLECULE TYPE:  protein
SQ   SEQUENCE   334 AA; 37627 MW; 566213 CN;

  Query Match           17.3%;  Score 69;  DB 3;  Length 334;
  Best Local Similarity  32.0%;  Pred. No. 7.04e+01;
  Matches     8;  Conservative    9;  Mismatches   7;  Indels   1;  Gaps   1;

Db     290 AGRVIDNPREALSAAMEEAQAWRKK 314
           ||::  |: : || :|: :  ||::
Qy      29 AGQIYGNSNQ-LSLSMDSPVEWRRR 52


RESULT    7
ID   US-08-472-595-9      STANDARD;      PRT:    334 AA.
XX
AC   XXXXXX
XX
DI
XX
DE   Sequence 9, Application US/08472595
XX
CC    Sequence 9, Application US/08472595
CC    Patent No. 6001583
CC     GENERAL INFORMATION:
CC       APPLICANT:  Marqolis, Benjamin L.
CC       TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR TREATMENT
CC       TITLE OF INVENTION:  OF BREAST CANCER
CC       NUMBER OF SEQUENCES:  20
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  PENNIE & EDMONDS LLP
CC         STREET:  1155 Avenue of the Americas
CC         CITY:  New York
CC         STATE:  New York
CC         COUNTRY:  U.S.A.
CC         ZIP:  10036-2711
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/472,595
CC         FILING DATE:  06-JUN-1995
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Coruzzi, Laura A.
CC         REGISTRATION NUMBER:  30,742
CC         REFERENCE/DOCKET NUMBER:  7684-103
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (212) 790-9090
CC         TELEFAX:  (212) 869-9741/8864
CC         TELEX:  66141 PENNIE
CC     INFORMATION FOR SEQ ID NO:  9:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  334 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  unknown
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   334 AA; 37627 MW; 566213 CN;

  Query Match           17.3%;  Score 69;  DB 3;  Length 334;
  Best Local Similarity  32.0%;  Pred. No. 7.04e+01;
  Matches     8;  Conservative    9;  Mismatches   7;  Indels   1;  Gaps   1;

Db     290 AGRVIDNPREALSAAMEKAQAWRKK 314
           ||::  |: : || :|: :  ||::
Qy      29 AGQIYGNSNQ LSLSMDSPVEWRRR 52


RESULT    8
ID   US-08-866-381A-1      STANDARD;      PRT:    335 AA.
XX
AC   XXXXXX
XX
DT
XX
DE   Sequence 1, Application US/08866381A
XX
CC    Sequence 1, Application US/08866381A
CC    Patent No. 6045797
CC     GENERAL INFORMATION:
CC       APPLICANT:  Ben Lewis Marqolis
CC       APPLICANT:  Joseph Schlessinger
CC       TITLE OF INVENTION:  METHODS FOR TREATMENT OR DIAGNOSIS
CC       TITLE OF INVENTION:  OF DISEASES OR CONDITIONS ASSOCIATED
CC       TITLE OF INVENTION:  WITH A BLM DOMAIN
CC       NUMBER OF SEQUENCES:  6
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Lyon & Lyon
CC         STREET:  633 West Fifth Street
CC         STREET:  Suite 4700
CC         CITY:  Los Angeles
CC         STATE:  California
CC         COUNTRY:  U.S.A.
CC         ZIP:  90071-2066
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
CC         MEDIUM TYPE:  storage
CC         COMPUTER:  IBM Compatible
CC         OPERATING SYSTEM:  IBM P.C. DOS 5.0
CC         SOFTWARE:  FastSEQ for Windows 2.0
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/866,381A
CC         FILING DATE:  May 30, 1997
CC         CLASSIFICATION:  530
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  08/212,234
CC         FILING DATE:  March 14, 1994
CC         APPLICATION NUMBER:
CC         FILING DATE:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Warburg, Richard J.
CC         REGISTRATION NUMBER:  32,327
CC         REFERENCE/DOCKET NUMBER:  22b/043
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (213) 489-1600
CC         TELEFAX:  (213) 955-0440
CC         TELEX:  67-3510
CC     INFORMATION FOR SEQ ID NO:  1:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  335 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
```

```
CC       FEATURE:
CC         OTHER INFORMATION:  BLM domain of GRB-7
SQ   SEQUENCE   335 AA; 37724 MW; 569630 CN;

  Query Match           17.4%;  Score 69;  DB 3;  Length 335;
  Best Local Similarity  42.0%;  Pred. No. 7.04e+01;
  Matches    8;  Conservative    9;  Mismatches    7;  Indels    1;  Gaps    1;

Db     291 AGRVIDNPREALSAAMEEAQAWRKK 315
           ||::  |: : || :|: :  |||:
Qy      29 AGQIYGNSNQ-LSLSMDSPVEWRRR 52


RESULT    9
ID   US-08-866-381A-5      STANDARD;      PRT;   534 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 5, Application US/08866381A
XX
CC    Sequence 5, Application US/08866381A
CC    Patent No. 6045797
CC     GENERAL INFORMATION:
CC       APPLICANT:  Ben Lewis Margolis
CC       APPLICANT:  Joseph Schlessinger
CC       TITLE OF INVENTION:  METHODS FOR TREATMENT OR DIAGNOSIS
CC       TITLE OF INVENTION:  OF DISEASES OR CONDITIONS ASSOCIATED
CC       TITLE OF INVENTION:  WITH A BLM DOMAIN
CC       NUMBER OF SEQUENCES:  6
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Lyon & Lyon
CC         STREET:  633 West Fifth Street
CC         STREET:  Suite 4700
CC         CITY:  Los Angeles
CC         STATE:  California
CC         COUNTRY:  U.S.A.
CC         ZIP:  90071-2066
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
CC         MEDIUM TYPE:  storage
CC         COMPUTER:  IBM Compatible
CC         OPERATING SYSTEM:  IBM P.C. DOS 5.0
CC         SOFTWARE:  FastSEQ for Windows 2.0
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/866,381A
CC         FILING DATE:  May 30, 1997
CC         CLASSIFICATION:  530
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  08/212,234
CC         FILING DATE:  March 14, 1994
CC         APPLICATION NUMBER:
CC         FILING DATE:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Warburg, Richard J.
CC         REGISTRATION NUMBER:  32,327
CC         REFERENCE/DOCKET NUMBER:  225/043
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (213) 489-1600
CC         TELEFAX:  (213) 955-0440
CC         TELEX:  67-3510
CC     INFORMATION FOR SEQ ID NO:  5:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  534 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
CC       FEATURE:
CC         OTHER INFORMATION:  GRB-7
SQ   SEQUENCE   534 AA; 59827 MW; 1432854 CN;

  Query Match           17.4%;  Score 69;  DB 4;  Length
  Best Local Similarity  42.0%;  Pred. No. 7.04e+01;
  Matches    8;  Conservative    9;  Mismatches    7;

Db     383 AGRVIDNPREALSAAMEEAQAWRKK 407
           ||::  |: : || :|: :  |||:
Qy      29 AGQIYGNSNQ-LSLSMDSPVEWRRR 52


RESULT   10
ID   US 07 906 349A 10      STANDARD;      PRT;   535 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 10, Application US/07906349A
XX
CC    Sequence 10, Application US/07906349A
CC    Patent No. 5434064
CC     GENERAL INFORMATION:
CC       APPLICANT:  Schlessinger, Joseph
CC       APPLICANT:  Skolnik, Edward Y.
CC       APPLICANT:  Margolis, Benjamin L.
CC       TITLE OF INVENTION:  A NOVEL EXPRESSION-CLONIN
CC       TITLE OF INVENTION:  IDENTIFYING TARGET PROTEIN
CC       TITLE OF INVENTION:  TARGET PROTEINS
CC       NUMBER OF SEQUENCES:  16
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Browdy and Neimark
CC         STREET:  419 Seventh Street, N.W.
CC         CITY:  Washington
CC         STATE:  D.C.
CC         COUNTRY:  USA
CC         ZIP:  20004
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/07/906,349A
CC         FILING DATE:  30-JUN-1992
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  07/764,237
CC         FILING DATE:  18-JAN-1991
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  202-628-5197
CC         TELEFAX:  202-737-3528
CC     INFORMATION FOR SEQ ID NO:  10:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  535 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   535 AA; 59959 MW; 1432872 CN;

  Query Match           17.4%;  Score 69;  DB 1;  Length
  Best Local Similarity  42.0%;  Pred. No. 7.04e+01;
  Matches    8;  Conservative    9;  Mismatches    7;

Db     384 AGRVIDNPREALSAAMEEAQAWRKK 408
           ||::  |: : || :|: :  |||:
Qy      29 AGQIYGNSNQ-LSLSMDSPVEWRRR 52


RESULT   11
ID   US-08-167-035-10      STANDARD;      PRT;   535 AA.
XX
AC   xxxxxx
XX
```

```
DT
XX
DE   Sequence 10, Application US/08167035
XX
CC    Sequence 10, Application US/08167035
CC    Patent No. 5618691
CC     GENERAL INFORMATION:
CC       APPLICANT:  Schlessinger, Joseph
CC       APPLICANT:  Skolnick, Edward Y.
CC       APPLICANT:  Margolis, Benjamin L.
CC       TITLE OF INVENTION:  NOVEL EXPRESSION CLONING METHOD FOR
CC       TITLE OF INVENTION:  IDENTIFYING TARGET PROTEINS FOR EUKARYOTIC TYROSINE
CC       TITLE OF INVENTION:  KINASES AND NOVEL TARGET PROTEINS
CC       NUMBER OF SEQUENCES:  50
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  PENNIE & EDMONDS
CC         STREET:  1155 Avenue of the Americas
CC         CITY:  New York
CC         STATE:  New York
CC         COUNTRY:  10036-2711
CC         ZIP:  10036-2711
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/167,035
CC         FILING DATE:  16-DEC-1993
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Coruzzi, Laura A.
CC         REGISTRATION NUMBER:  30,742
CC         REFERENCE/DOCKET NUMBER:  7683-062
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (212) 790-9090
CC         TELEFAX:  (212) 869 9741/8864
CC         TELEX:  66141 PENNIE
CC     INFORMATION FOR SEQ ID NO:  10:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  535 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  unknown
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   535 AA; 59973 MW; 1438700 CN;

  Query Match           17.3%;  Score 69;  DB 1;  Length 535;
  Best Local Similarity  32.0%;  Pred. No. 7.04e+01;
  Matches    8;  Conservative    9;  Mismatches   7;  Indels   1;  Gaps   1;

Db     384 AGRVIDNPREALSAAMEEAQAWRKK 408
           ||::  |: : || :|: :  ||::
Qy      29 AGQIYGNSNQ-LSLSMDSPVEWRRR 52


RESULT   12
ID   US-08-539-005-10     STANDARD;     PRT;   535 AA.
XX
AC   XXXXXX
XX
DT
XX
DE   Sequence 10, Application US/08539005
XX
CC    Sequence 10, Application US/08539005
CC    Patent No. 5858686
CC     GENERAL INFORMATION:
CC       APPLICANT:  Schlessinger, Joseph
CC       APPLICANT:  Skolnick, Edward Y.
CC       APPLICANT:  Margolis, Benjamin L.
CC       TITLE OF INVENTION:  NOVEL EXPRESSION CLONING METHOD FOR
CC       TITLE OF INVENTION:  IDENTIFYING TARGET PROTEINS FOR EUKARYOTIC TYROSINE
CC       TITLE OF INVENTION:  KINASES AND NOVEL TARGET PROTEINS
CC       NUMBER OF SEQUENCES:  50
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  PENNIE & EDMONDS
CC         STREET:  1155 Avenue of the Americas
CC         CITY:  New York
CC         STATE:  New York
CC         COUNTRY:  10036-2711
CC         ZIP:  10036-2711
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/539,005
CC         FILING DATE:  4-OCT-1995
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 08/167,035
CC         FILING DATE:  16-DEC-1993
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Coruzzi, Laura A.
CC         REGISTRATION NUMBER:  30,742
CC         REFERENCE/DOCKET NUMBER:  7683-062
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (212) 790-9090
CC         TELEFAX:  (212) 869-9741/8864
CC         TELEX:  66141 PENNIE
CC     INFORMATION FOR SEQ ID NO:  10:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  535 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  unknown
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   535 AA; 59973 MW; 1438700 CN;

  Query Match           17.3%;  Score 69;  DB 2;  Length 535;
  Best Local Similarity  32.0%;  Pred. No. 7.04e+01;
  Matches    8;  Conservative    9;  Mismatches   7;  Indels   1;  Gaps   1;

Db     384 AGRVIDNPREALSAAMEEAQAWRKK 408
           ||::  |: : || :|: :  ||::
Qy      29 AGQIYGNSNQ-LSLSMDSPVEWRRR 52


RESULT   13
ID   US-08-208-887A-10     STANDARD;     PRT;   535 AA.
XX
AC   XXXXXX
XX
DT
XX
DE   Sequence 10, Application US/08208887A
XX
CC    Sequence 10, Application US/08208887A
CC    Patent No. 5677421
CC     GENERAL INFORMATION:
CC       APPLICANT:  Schlessinger, Joseph
CC       APPLICANT:  Skolnick, Edward Y.
CC       APPLICANT:  Margolis, Benjamin L.
CC       TITLE OF INVENTION:  NOVEL EXPRESSION CLONING METHOD FOR
CC       TITLE OF INVENTION:  IDENTIFYING TARGET PROTEINS FOR EUKARYOTIC TYROSINE
CC       TITLE OF INVENTION:  KINASES AND NOVEL TARGET PROTEINS
CC       NUMBER OF SEQUENCES:  51
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  PENNIE & EDMONDS
CC         STREET:  1155 Avenue of the Americas
CC         CITY:  New York
CC         STATE:  New York
CC         COUNTRY:  10036-2711
CC         ZIP:  10036-2711
CC       COMPUTER READABLE FORM:
```

CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/208,887A
CC FILING DATE: 11-MAR-1994
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Coruzzi, Laura A.
CC REGISTRATION NUMBER: 30,742
CC REFERENCE/DOCKET NUMBER: 7683-063
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 790-9090
CC TELEFAX: (212) 869 9741/8864
CC TELEX: 66141 PENNIE
CC INFORMATION FOR SEQ ID NO: 10:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 535 amino acids
CC TYPE: amino acid
CC TOPOLOGY: unknown
CC MOLECULE TYPE: protein
SQ SEQUENCE 535 AA; 59973 MW; 1438700 CN;

Query Match 17.3%; Score 69; DB 1; Length 535;
Best Local Similarity 32.0%; Pred. No. 7.04e+01;
Matches 8; Conservative 9; Mismatches 7; Indels 1; Gaps 1;

Db 384 AGRVIDNPREALSAAMEEAQAWRRK 408
Qy 29 AGQIYGNSNQ-LSLSMQSPVEWRRR 52

RESULT 14
ID US-08-121-713D-60 STANDARD; PRT: 650 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 60, Application US/08121713D
XX
CC Sequence 60, Application US/08121713D
CC Patent No. 5639856
CC GENERAL INFORMATION:
CC APPLICANT: Goodman, Corey S.
CC APPLICANT: Kolodkin, Alex L.
CC APPLICANT: Matthes, David
CC APPLICANT: Bentley, David R.
CC APPLICANT: O'Connor, Timothy
CC TITLE OF INVENTION: The Semaphorin Gene Family
CC NUMBER OF SEQUENCES: 100
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
CC STREET: 268 Bush Street, Suite 3200
CC CITY: San Francisco
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94104
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/121,713D
CC FILING DATE: 13-SEP-1993
CC CLASSIFICATION: 514
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Osman, Richard A.
CC REGISTRATION NUMBER: 36,627
CC REFERENCE/DOCKET NUMBER: B94-002-1
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415)543 4543
CC TELEFAX: (415) 543-4543
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 60:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 650 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 650 AA; 72940 MW; 2193467 CN;

Query Match 17.3%; Score 69; DB 1; Length
Best Local Similarity 32.0%; Pred. No. 7.04e+
Matches 8; Conservative 9; Mismatches

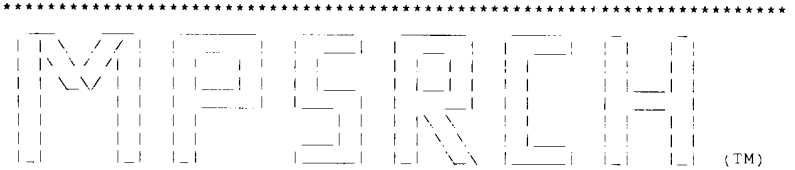
Db 46 TKNGQAVCPYDPRHNSISVLAQNLLYSGIVA-DPSGSDPTIYR
Qy 29 AGQIYGNSNQ-LSLSMQSPVEWRRR

RESULT 15
ID US-08-833-391-60 STANDARD; PRT: 650 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 60, Application US/08833391
XX
CC Sequence 60, Application US/08833391
CC Patent No. 6013781
CC GENERAL INFORMATION:
CC APPLICANT: Goodman, Corey S.
CC APPLICANT: Kolodkin, Alex L.
CC APPLICANT: Matthes, David
CC APPLICANT: Bentley, David R.
CC APPLICANT: O'Connor, Timothy
CC TITLE OF INVENTION: The Semaphorin Gene Fami
CC NUMBER OF SEQUENCES: 100
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
CC STREET: 268 Bush Street, Suite 3200
CC CITY: San Francisco
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94104
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/833,391
CC FILING DATE:
CC CLASSIFICATION: 530
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/121,713
CC FILING DATE: 13 SEP 1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Osman, Richard A.
CC REGISTRATION NUMBER: 36,627
CC REFERENCE/DOCKET NUMBER: B94-002-1
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415)543 4543
CC TELEFAX: (415) 543-4543
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 60:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 650 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 650 AA; 72940 MW; 2193467 CN;

Query Match 17.4%; Score 69; DB 3; Length 650;
Best Local Similarity 20.9%; Pred. No. 7.04e+01;
Matches 9; Conservative 13; Mismatches 20; Indels 1; Gaps 1;

```
Db      46 IKNCQAVCPYDPRHNSISVLADNELYSGTVA-DFSGSDPIIYR 87
           ::|:  || |    :: ::|   ::|: :    :|   |:  |
Qy       8 SRNSSRVCNYHGCSSAICIIFAGQIYGNSNQLSLSMDSPVEWR 50
```

Search completed: Fri Aug 25 18:09:37 2000
Job time : 9 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 25 18:07:21 2000; MasPar time 5.71 Seconds
446.355 Million cell updates/sec

Tabular output not generated.

Title: >US-09-257-585-8
Description: (1-54) from US09257585.pep
Perfect Score: 398
Sequence: 1 MVIEGYLSRNSSKVCNYBGC..... NSNQLSLSMDSPVEWRRRI 54

Scoring table: PAM 150
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir64
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 35.776; Variance 58.217; scale 0.615

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	94	23.6	496	2	A56040	protein-tyrosine kina	3.43e-03
2	94	23.6	496	2	I56322	srm - mouse	3.43e-03
3	90	22.6	314	2	T11711	probable small GTPase	1.65e-02
4	83	20.9	1550	2	S60220	glutamate synthase (f	2.34e-01
5	82	20.6	683	2	B72539	hypothetical protein	3.37e-01
6	79	19.9	212	2	S23934	lipase lipA - Bacillu	1.00e+00
7	79	19.8	411	2	B72781	hypothetical protein	1.00e+00
8	79	19.8	706	2	T06228	hemolysin B - Escheri	1.00e+00
9	78	19.6	434	2	I01390	hypothetical protein	1.43e+00
10	78	19.6	582	2	A70841	probable oxalyl-CoA d	1.43e+00
11	77	19.3	81	2	B22848	hypothetical ORF-3 pr	2.03e+00
12	77	19.3	92	2	D25877	MURF3 protein - Crith	2.03e+00
13	76	19.1	278	1	EDBE63	immediate-early prote	2.88e+00
14	76	19.1	333	1	I52525	testin precursor - ra	2.88e+00
15	76	19.1	428	2	C64725	fixC protein - Escher	2.88e+00
16	76	19.1	758	2	S30948	ferrichrome receptor	2.88e+00
17	75	18.8	425	2	H70902	GTP cyclohydrolase II	1.07e+00
18	75	18.8	1256	2	S60461	gene flightless I pro	4.07e+00
19	74	18.6	266	2	A39260	MHC class II histocom	5.74e+00
20	74	18.6	266	2	B39260	MHC class II histocom	5.74e+00
21	74	18.6	664	2	PC4002	phosphatidylinositol-	5.74e+00
22	74	18.6	725	2	I08989	hypothetical protein	5.74e+00
23	74	18.6	808	2	B70747	probable uridylyltran	5.74e+00
24	73	18.3	549	2	S23477	hypothetic	[illegible]
25	73	18.3	656	2	S09611	hypothetic	[illegible]
26	73	18.3	1133	2	I60192	probable K+	[illegible]
27	73	18.3	1138	2	I09920	hypothetic	[illegible]
28	72	18.1	217	2	B25697	H-2 class	[illegible]
29	72	18.1	227	2	JQ0140	hypothetic	[illegible]
30	72	18.1	5147	1	CJHUFM	cadherin-1	[illegible]
31	71	17.8	214	1	A61809	vit protei	[illegible]
32	71	17.8	227	2	B70672	probable [illegible]	[illegible]
33	71	17.8	509	2	S38853	methionyl	[illegible]
34	71	17.8	519	2	S42903	phospholip	[illegible]
35	71	17.8	521	2	B72384	bacteriofe	[illegible]
36	71	17.8	543	2	B44491	hypothetic	[illegible]
37	71	17.8	655	2	S31162	probable t	[illegible]
38	71	17.8	809	2	A41603	nitrogenas	[illegible]
39	71	17.8	859	2	S54473	phosphoqly	[illegible]
40	70	17.6	256	2	B72078	nadh (ubic	[illegible]
41	70	17.6	486	2	A56044	NADH dehyd	[illegible]
42	70	17.6	542	2	S27856	long chain	[illegible]
43	70	17.6	827	1	A60HY1	KEDI prote	[illegible]
44	70	17.6	957	2	B42709	exonuclea	[illegible]
45	70	17.6	1961	2	S50420	sodium cha	[illegible]

ALIGNMENTS

RESULT 1
ENTRY A56040 #type complete
TITLE protein-tyrosine kinase (EC 2.7.1.112) - mouse
ORGANISM #formal_name Mus musculus #common_name
DATE 01-Dec-1995 #sequence_revision 01-Dec-1995 #text_change 24-Sep-1999
ACCESSIONS A56040
REFERENCE A56040
#authors Kohmura, N.; Yuri, K.; Tomooka, Y.; Oji, R.; Takeda, N.; Chiba, J.; Ikawa, Y.
#journal Mol. Cell. Biol. (1994) 14:6915-6925
#title A novel nonreceptor tyrosine kinase, s... targeted disruption.
#cross-references MUID:95022122
#accession A56040
##status preliminary
##molecule_type mRNA
##residues 1-496 ##label KOH
##cross-references GB:D27064; NID:g520072; PIDN:... PID:d1007...; PID:g529073
GENETICS
#map_position 2
CLASSIFICATION #superfamily unassigned Ser/Thr or Tyr... kinases, protein kinase homology; SH... homology
KEYWORDS ATP; phosphotransferase
FEATURE
62-111 #domain SH3 homology #label SH3
124-216 #domain SH2 homology #label SH2
232-491 #domain protein kinase homology #...
240-248 #region protein kinase ATP-bindi...
SUMMARY #length 496 #molecular-weight 55593

Query Match 23.6%; Score 94; DB 2; Len...
Best Local Similarity 46.6%; Pred. No. 3.43e-03;
Matches 19; Conservative 12; Mismatches 10; ...

Db 157 [illegible]
Qy 3 [illegible]

RESULT 2
ENTRY I56322 #type complete
TITLE srm - mouse
ORGANISM #formal_name Mus musculus #common_name

```
DATE            26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change
                  24-Sep-1999
ACCESSIONS      I56322
REFERENCE       I56322
   #authors     Kawachi, Y.; Nakauchi, H.; Otsuka, F.
   #journal     J. Invest. Dermatol. (1995) 21:533-538
   #title       Identification of a novel cDNA clone encoding protein
                  tyrosine kinase in murine skin.
   #accession   I56322
      ##status       preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues     1-496 ##label RES
      ##cross-references GB:D49427; NID:g684971; PIDN:BAA08406.1; PID:g684972
CLASSIFICATION  #superfamily unassigned Ser/Thr or Tyr-specific protein
                  kinases; protein kinase homology; SH2 homology; SH3
                  homology
KEYWORDS        ATP
FEATURE
   62-111            #domain SH3 homology #label SH3\
   124-216           #domain SH2 homology #label SH2\
   232-491           #domain protein kinase homology #label KIN\
   240 248           #region protein kinase ATP-binding motif
SUMMARY         #length 496  #molecular-weight 55731  #checksum 859

  Query Match             23.6%;  Score 94;  DB 2;  Length 496;
  Best Local Similarity   36.6%;  Pred. No. 3.43e-03;
  Matches    15;  Conservative    12;  Mismatches  10,  Indels    4,  Gaps    4,

Db    157 IGGYSLSVRAQAKVCHYRICMAPSGSLYLQEGQLFPSLDAL 197
          | || || |  ::||:|: | :: | : |: ||:: : : |
Qy      3 IEGY-LS-RNSSRVCNYHGCSSAIG-I-LEAGQIYGNSNQL 39


RESULT    3
ENTRY           T11711      #type complete
TITLE           probable small GTPase - fission yeast (Schizosaccharomyces
                  pombe)
ORGANISM        #formal_name Schizosaccharomyces pombe
DATE            16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change
                  16-Jul-1999
ACCESSIONS      T11711
REFERENCE       Z17318
   #authors     Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell,
                  B.G.
   #submission  submitted to the EMBL Data Library, August 1998
   #accession   T11711
      ##status       preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues     1-314 ##label SEE
      ##cross-references EMBL:AL031532; NID:e1319424; PID:e1319429
      ##experimental_source strain 972h(-)
GENETICS
   #map_position IIIR
   #introns      47,2
   #note         SPCC777.05
SUMMARY         #length 314  #molecular-weight 35591  #checksum 4744

  Query Match             22.6%;  Score 90;  DB 2;  Length 314;
  Best Local Similarity   45.9%;  Pred. No. 1.65e-02;
  Matches    17;  Conservative     6;  Mismatches  11;  Indels    3,  Gaps    3;

Db    209 YVATDSSPVDVQSYEICSDFIDVILDIGSIYGRSSQL 245
          |:: :|| |   :|  ||  |  ||: | ||| |:||
Qy      6 YLSRNSSRV-C-NYHGCSSAIG-ILEAGQIYGNSNQL 39


RESULT    4
ENTRY           S60228      #type complete
TITLE           glutamate synthase (ferredoxin) (EC 1.4.7.1) gltB -
                  Synechocystis sp. (strain PCC 6803)
ALTERNATE_NAMES protein sll1502
ORGANISM        #formal_name Synechocystis sp.
   #variety     PCC 6803

DATE            09-Mar-1996 #sequence_revision 12-Apr-1996 #text_change
                  11-Jun-1999
ACCESSIONS      S60228; S74978; S46957
REFERENCE       S60228
   #authors     Navarro, F.; Chavez, S.; Candau, P.; Florencio, F.J.
   #journal     Plant Mol. Biol. (1995) 27:753-767
   #title       Existence of two ferredoxin-glutamate synthases in the
                  cyanobacterium Synechocystis sp. PCC 6803: isolation and
                  insertional inactivation of gltB and gltS genes.
   #cross-references MUID:95244836
   #accession   S60228
      ##molecule_type DNA
      ##residues     1-1550 ##label NAV
      ##cross-references EMBL:X80485; NID:g515937; PIDN:CAA56652.1;
                        PID:g515938
      ##experimental_source Synechocystis sp. PCC 6803
      ##note        it is uncertain whether Met-1 or Met-21 is the initiator
REFERENCE       S74322
   #authors     Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
                  Nakamura, Y.; Miyajima, N.; Hirosawa, M.; Sugiura, M.;
                  Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
                  Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo,
                  S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
                  Yasuda, M.; Tabata, S.
   #journal     DNA Res. (1996) 3:109-136
   #title       Sequence analysis of the genome of the unicellular
                  cyanobacterium Synechocystis sp. PCC6803. II. Sequence
                  determination of the entire genome and assignment of
                  potential protein-coding regions.
   #cross-references MUID:97061201
   #accession   S74978
      ##status       nucleic acid sequence not shown; translation not shown
      ##molecule_type DNA
      ##residues     1-1550 ##label KAN
      ##cross-references EMBL:D90902; GB:AB001339; NID:g1652027;
                        PIDN:BAA17018.1; PID:d1017751; PID:g1652093
      ##experimental_source Synechocystis sp. PCC 6803
      ##note        the nucleotide sequence was submitted to the EMBL Data
                     Library, June 1996
GENETICS
   #gene        gltB
CLASSIFICATION  #superfamily glutamate synthase (NADPH)
KEYWORDS        3Fe-4S; flavoprotein; glutamate biosynthesis; iron-sulfur
                  protein; metalloprotein; oxidoreductase
FEATURE
   1150,1156,1161    #binding_site 3Fe-4S cluster (Cys) (covalent) #status
                       predicted
SUMMARY         #length 1550  #molecular-weight 169070  #checksum 6525

  Query Match             20.9%;  Score 83;  DB 2;  Length 1550;
  Best Local Similarity   26.4%;  Pred. No. 2.44e-01;
  Matches    14;  Conservative    15;  Mismatches  21;  Indels    3;  Gaps    3;

Db    777 VIDEYFCNTSSNIQGSDLGVIAQEAILRHQHAFAPPGDLHIIDVGGEYQWRK 829
          ||: |: | |||:     |  :  :||   : ::  : :| :| : :  :||:
Qy      2 VTEGYLSRNSSRVCNYH GCSSAIGILEAGQIYG-NSNQL-SLSMDSPVEWRR 51


RESULT    5
ENTRY           B72539      #type complete
TITLE           hypothetical protein APE1599 - Aeropyrum pernix (strain K1)
ORGANISM        #formal_name Aeropyrum pernix
DATE            20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change
                  20-Aug-1999
ACCESSIONS      B72539
REFERENCE       A72450
   #authors     Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.;
                  Haikawa, Y.; Jin-no, K.; Takahashi, M.; Sekine, M.; Baba,
                  S.; Ankai, A.; Kosugi, H.; Hosoyama, A.; Fukui, S.; Nagai,
                  Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
                  Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.;
                  Kushida, N.; Oguchi, A.; Aoki, K.; Kubota, K.; Nakamura,
                  Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
```

```
   #journal        DNA Res. (1999) 6:83 101
   #title          Complete genome sequence of an aerobic hyper-thermophilic
                     Crenarchaeon, Aeropyrum pernix K1.
   #cross-references MUID:99310339
   #accession      B72539
      ##status          preliminary
      ##molecule_type DNA
      ##residues       1-683 ##label KAW
      ##cross-references DDBJ:AP000062; NID:g5105244; PID:BAA80599.1;
                        PID:d1044385; PID:g5105286
      ##experimental_source strain K1
GENETICS
   #gene           APE1599
SUMMARY            #length 683  #molecular-weight 75228  #checksum 8462

  Query Match             20.6%;  Score 82;  DB 2;  Length 683;
  Best Local Similarity  36.4%;  Pred. No. 3.37e-01;
  Matches    12;  Conservative    8;  Mismatches  12;  Indels    1;  Gaps    1;

Db     631 SEVQVIRIGGYVILVSSEVLRLAEEMKAEWPRRR 668
           | : :|  :| :  :|: | |: :   ||||||
Qy      22 SAIGILE-AGQIYGNSNQLSLSMDSPVEWRRRR 54


RESULT    6
ENTRY             S23934      #type complete
TITLE             lipase lipA - Bacillus subtilis
ORGANISM          #formal_name Bacillus subtilis
DATE              22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change
                    02-Jul-1998
ACCESSIONS        S23934; B69652
REFERENCE         S23934
   #authors       Dartois, V.; Baulard, A.; Schanck, K.; Colson, C.
   #journal       Biochim. Biophys. Acta (1992) 1131:253-260
   #title         Cloning, nucleotide sequence and expression in Escherichia
                    coli of a lipase gene from Bacillus subtilis 168.
   #cross-references MUID:92329538
   #accession     S23934
      ##status          preliminary
      ##molecule_type DNA
      ##residues       1-212 ##label DAR
      ##cross-references GB:M74010; NID:g143153; PID:g143154
REFERENCE         A69580
   #authors       Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
                    Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
                    Bolotin, A.; Borchert, S.; Borriss, R.; Boursier, L.; Brans,
                    A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
                    Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
                    Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
                    Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
                    Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.;
                    Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
                    M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
                    S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
                    Guiseppi, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood,
                    C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
                    Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
                    Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
                    Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.;
                    Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
                    Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
                    Mauel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
                    M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
                    M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
                    V.; Pohl, T.M.; Portetelle, D.; Porwollik, S.; Prescott,
                    A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
                    Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
                    Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
                    Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
                    Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
                    B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
                    Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
                    Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
                    Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
                    Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.;
                    Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
                    K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
                    Yoshikawa, H.; Danchin, A.
   #journal       Nature (1997) 390:249-256
   #title         The complete genome sequence of the Gram-positive
                    Bacillus subtilis.
   #cross-references MUID:98044555
   #accession     B69652
      ##status          preliminary; nucleic acid sequence
                         translation not shown
      ##molecule_type DNA
      ##residues       1-192,'E',194-212 ##label KUN
      ##cross-references GB:Z99105; GB:AL009126; NID:g2
                        PID:g2632550
      ##experimental_source strain 168
GENETICS
   #gene           lipA
SUMMARY            #length 212  #molecular-weight 22777

  Query Match             20.6%;  Score 79;  DB 2;  Length
  Best Local Similarity  35.4%;  Pred. No. 1.00e+00;
  Matches    12;  Conservative    7;  Mismatches    9;

Db     165 MIVMNYLSRLDGARNVQIHRVAE IGILYSSQV 196
           |::  |||| :::|    || :  :|:: :::|
Qy       1 MVIEGYLSR-NSSKVCNYHCTSSAIGILEAGQI 32


RESULT    7
ENTRY             B72781      #type complete
TITLE             hypothetical protein APE0235 - Aeropyrum
ORGANISM          #formal_name Aeropyrum pernix
DATE              20-Aug-1999 #sequence_revision 20-Aug
                    20-Aug-1999
ACCESSIONS        B72781
REFERENCE         A72450
   #authors       Kawarabayasi, Y.; Hino, Y.; Horikawa,
                    Haikawa, Y.; Jin-no, K.; Takahashi,
                    S.; Ankai, A.; Kosugi, H.; Hosoyama
                    Y.; Nishijima, K.; Nakazawa, H.; Tan
                    Funahashi, T.; Tanaka, T.; Kudoh, Y
                    Kushida, N.; Harasawa, A.; Aoki, K.; F
                    Y.; Nomura, N.; Sako, Y.; Kikuchi,
   #journal       DNA Res. (1999) 6:83-101
   #title         Complete genome sequence of an aerobi
                    Crenarchaeon, Aeropyrum pernix K1.
   #cross-references MUID:99310339
   #accession     B72781
      ##status          preliminary
      ##molecule_type DNA
      ##residues       1-311 ##label KAW
      ##cross-references DDBJ:AP0[illegible]; NID:g5105[illegible]
                        PID:d1042924; PID:g5105627
      ##experimental_source strain K1
GENETICS
   #gene           APE0235
SUMMARY            #length 311  #molecular-weight 33221

  Query Match             19.8%;  Score 79;  DB 2;  Length
  Best Local Similarity  27.5%;  Pred. No. 1.00e+00;
  Matches    14;  Conservative   13;  Mismatches   13;

Db      16 LSSEAGSEACWGYIAVAVGLAAIGSMRSSRLF WHSIAG
           || :: |: |  | :::: :| : : :: :|:|: | :
Qy       7 LSRNSSKVCNYHCTSSAIGILEAGQIGNQLSLSMDSPVE

RESULT    8
ENTRY             I00228      #type complete
TITLE             hemolysin B - Escherichia coli plasmi
ORGANISM          #formal_name Escherichia coli
```

```
DATE              01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change
                    20-Aug-1999
ACCESSIONS        T00228
REFERENCE         Z14127
   #authors       Makino, K.; Ishii, K., Yasunaga, T., Hattori, M., Yokoyama,
                    K.; Yatsudo, H.C.; Kubota, Y.; Yamaichi, Y.; Iida, T.;
                    Yamamoto, K., Honda, T., Han, C., Ohtsubo, A.; Kasamatsu,
                    M.; Hayashi, T.; Kuhara, S.; Shinagawa, H.
   #journal       DNA Res. (1998) 5:1-9
   #title         Complete nucleotide sequences of 93 kb and 3.3-kb plasmids of
                    an enterohemorrhagic Escherichia coli O157:H7 derived from
                    Sakai outbreak.
   #cross-references MUID:98290540
   #accession     T00228
      ##status         preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues       1-706 ##label MAK
      ##cross-references EMBL:AB011549; NID:d1204561; PIDN:BAA31775.1;
                        PID:d1032736
      ##experimental_source strain EHEC O157:H7, substrain RIMD 0509952
GENETICS
   #gene          hlyB
   #genome        plasmid pO157
CLASSIFICATION    #superfamily hemolysin secretion protein B; ATP binding
                    cassette homology
KEYWORDS          ATP
FEATURE
   485-679             #domain ATP-binding cassette homology #label ABC
SUMMARY           #length 706  #molecular-weight 79814  #checksum 899

  Query Match           19.8%;  Score 79;  DB 2;  Length 706;
  Best Local Similarity  51.7%;  Pred. No. 1.00e+00;
  Matches   15;  Conservative    7;  Mismatches   1;  Indels   3;  Gaps   3;

Db    520 IPETGQILIDGHDLSLA-D-P-EWLRRQI 545
          | |:|||  ::::|||: | | || ||:|
Qy     26 ILEAGQIYGNSNQLSLSMDSPVEWRRRI 54


RESULT    9
ENTRY             T01390      #type complete
TITLE             hypothetical protein T4I9.9 - Arabidopsis thaliana
ALTERNATE_NAMES   hypothetical protein OR23
ORGANISM          #formal_name Arabidopsis thaliana #common_name mouse-ear
                    cress
DATE              12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
                    24-Mar-1999
ACCESSIONS        T01390; S71216
REFERENCE         Z14314
   #authors       Parnell, L.D.; Gholj, L.; de la Bastide, M.; Hameed, A.;
                    Habermann, K.; Schutz, K., Huang, E.; Cattolico, L.;
                    Dedhia, N.N.; McCombie, W.R.
   #submission    submitted to the EMBL Data Library, May 1998
   #description   Genomic sequence of BAC T4I9 from Arabidopsis thaliana,
                    chromosome IV, near 16.6 cM
   #accession     T01390
      ##status         translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues       1-434 ##label PAR
      ##cross-references EMBL:AP000442; NID:g3242970; PID:g3924602
      ##experimental_source cultivar Columbia
REFERENCE         S71216
   #authors       Loebler, M.
   #submission    submitted to the EMBL Data Library, October 1995
   #description   OR23 is a 5' truncated cDNA sequence from Arabidopsis.
   #accession     S71216
      ##molecule_type mRNA
      ##residues       243-434 ##label LOE
      ##cross-references EMBL:U37704; NID:g1022808; PID:g1022809
GENETICS
   #map_position 4
   #introns       34/2; 76/2
   #note          T4I9.9

SUMMARY           #length 434  #molecular-weight 49373  #checksum 9801

  Query Match           19.6%;  Score 78;  DB 2;  Length 434;
  Best Local Similarity  29.4%;  Pred. No. 1.43e+00;
  Matches   10;  Conservative    7;  Mismatches  17;  Indels   6;  Gaps   0;

Db    268 YGGSRTVSGILPVDEYYKDAVVMDLKVDGGEKWR 301
          | |  :  |||   : | ::  : | :|:   ||
Qy     17 YHGCSSAIGILEAGQIYGNSNQLSLSMDSPVEWR 50


RESULT   10
ENTRY             A70841      #type complete
TITLE             probable oxalyl-CoA decarboxylase  Mycobacterium
                    tuberculosis (strain H37RV)
ORGANISM          #formal_name Mycobacterium tuberculosis
DATE              17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
                    29-Sep-1999
ACCESSIONS        A70841
REFERENCE         A70500
   #authors       Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
                    C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry
                    III, C.E.; Tekaia, F.; Badcock, K.; Basham, D.; Brown, D.;
                    Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
                    Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
                    Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
                    Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
                    Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
                    Skelton, S.; Squares, S.; Squares, R.; Sulston, J.E.;
                    Taylor, K.; Whitehead, S.; Barrell, B.G.
   #journal       Nature (1998) 393:537 544
   #title         Deciphering the biology of Mycobacterium tuberculosis from
                    the complete genome sequence.
   #cross-references MUID:98295987
   #accession     A70841
      ##status         preliminary; nucleic acid sequence not shown;
                       translation not shown
      ##molecule_type DNA
      ##residues       1-582 ##label COL
      ##cross-references GB:AL021926; GB:AL123456; NID:g2261520;
                        PIDN:CAA17312.1; PID:e1252410; PID:g2909583
      ##experimental_source strain H37Rv
GENETICS
   #gene          oxcA
CLASSIFICATION    #superfamily oxalyl-CoA decarboxylase; thiamine
                    pyrophosphate binding domain homology
SUMMARY           #length 582  #molecular-weight 61171  #checksum 4243

  Query Match           19.6%;  Score 78;  DB 2;  Length 582;
  Best Local Similarity  23.4%;  Pred. No. 1.43e+00;
  Matches   11;  Conservative   14;  Mismatches  20;  Indels   2;  Gaps   2;

Db    457 GFSGMEFETICRYRL-PVTVVILNNGGVYRGDFATIPPSAAPV-WRH 501
          |       | |   . . ||: | :|   :   :   :|| ||:
Qy      5 GYLSPNSSPVCNYHGCSSAIGILEAGQIYGNSNQLSLSMDSPVEWRR 51


RESULT   11
ENTRY             E22848      #type fragment
TITLE             hypothetical ORF-3 protein  Leishmania tarentolae
                    mitochondrion (SGC6) (fragment)
ORGANISM          #formal_name mitochondrion Leishmania tarentolae
DATE              30-Jun-1989 #sequence_revision 12-Sep-1997 #text_change
                    21-Aug-1998
ACCESSIONS        E22848
REFERENCE         A22848
   #authors       de la Cruz, V.F.; Neckelmann, N.; Simpson, L.
   #journal       J. Biol. Chem. (1984) 259:15136-15147
   #title         Sequences of six genes and several open reading frames in the
                    kinetoplast maxicircle DNA of Leishmania tarentolae.
   #cross-references MUID:85079995
   #accession     E22848
      ##molecule_type DNA
```

```
      ##residues       1-81 ##label DEL
      ##cross-references GB:M10126
GENETICS
   #genome         mitochondrion
   #genetic_code SGC6
CLASSIFICATION   #superfamily NADH dehydrogenase (ubiquinone) 49K protein
KEYWORDS         mitochondrion
SUMMARY          #length 81  #checksum 8754

  Query Match           19.3%;  Score 77;  DB 2;  Length 81;
  Best Local Similarity  50.0%;  Pred. No. 2.03e+00;
  Matches    8;  Conservative    4;  Mismatches   4;  Indels   0;  Gaps   0;

Db      48 VIIGYLHRGTEKLCEY 63
           || ||| | : ::|:|
Qy       2 VIEGYLSRNSSRVCNY 17


RESULT   12
ENTRY            D25877      #type fragment
TITLE            MURF3 protein - Crithidia fasciculata mitochondrion (SGC6)
                   (fragment)
ORGANISM         #formal_name mitochondrion Crithidia fasciculata
DATE             19-Jan-1988 #sequence_revision 19-Jan-1988 #text_change
                   21-Aug-1998
ACCESSIONS       D25877; S01163
REFERENCE        A25877
   #authors      Sloof, P.; van den Burg, J.; Voogd, A.; Benne, R.
   #journal      Nucleic Acids Res. (1987) 15:51-65
   #title        The nucleotide sequence of a 3.2 kb segment of mitochondrial
                   maxicircle DNA from Crithidia fasciculata containing the
                   gene for cytochrome oxidase subunit III, the N-terminal
                   part of the apocytochrome b gene and a possible frameshift
                   gene; further evidence for the use of unusual initiator
                   triplets in trypanosome mitochondria.
   #cross-references MUID:87146364
   #accession    D25877
      ##status         preliminary; not compared with conceptual translation
      ##molecule_type DNA
      ##residues       1-92 ##label SLO
REFERENCE        S01163
   #authors      van der Spek, H.; van den Burg, J.; Croiset, A.; van den
                   Broek, M.; Sloof, P.; Benne, R.
   #journal      EMBO J. (1988) 7:2509-2514
   #title        Transcripts from the frameshifted MURF3 gene from Crithidia
                   fasciculata are edited by U insertion at multiple sites.
   #cross-references MUID:89052665
   #accession    S01163
      ##status         not compared with conceptual translation
      ##molecule_type mRNA
      ##residues       24-66 ##label VAN
GENETICS
   #gene           MURF3
   #genome         mitochondrion
   #genetic_code SGC6
CLASSIFICATION   #superfamily NADH dehydrogenase (ubiquinone) 49K protein
KEYWORDS         mitochondrion; RNA editing
SUMMARY          #length 92  #checksum 2667

  Query Match           19.3%;  Score 77;  DB 2;  Length 92;
  Best Local Similarity  50.0%;  Pred. No. 2.03e+00;
  Matches    8;  Conservative    4;  Mismatches   4;  Indels   0;  Gaps   0;

Db      63 VIIGYLHRGTEKLCEY 78
           || ||| | : ::|:|
Qy       2 VIEGYLSRNSSRVCNY 17


RESULT   13
ENTRY            EDHE63      #type complete
TITLE            immediate-early protein IE68 - human herpesvirus 3
ORGANISM         #formal_name human herpesvirus 3, varicella-zoster virus
DATE             30-Sep-1988 #sequence_revision 30-Sep-1988 #text_change
                   16-Jul-1999
ACCESSIONS       B27345; I27345
REFERENCE        A27345
   #authors      Davison, A.J.; Scott, J.E.
   #journal      J. Gen. Virol. (1986) 67:1759-1816
   #title        The complete DNA sequence of varicella-
   #cross-references MUID:86300057
   #accession    B27345
      ##molecule_type DNA
      ##residues       1-276 ##label VAV
      ##cross-references EMBL:X04370; NID:g59989; PIDN:
GENETICS
   #gene           63; 70
CLASSIFICATION   #superfamily herpesvirus immediate-ear
KEYWORDS         early protein
SUMMARY          #length 278  #molecular-weight 31455

  Query Match           19.1%;  Score 76;  DB 1;  Leng
  Best Local Similarity  48.8%;  Pred. No. 2.88e+00;
  Matches    8;  Conservative    3;  Mismatches   7;

Db     239 EIAE-YIAAKALIAGGCVIWKeeK 263
           |::: | :: |:   :: |: :::
Qy      28 EAGQIYGNSNQLSLSMESFVLWPeeK 53


RESULT   14
ENTRY            I52525      #type complete
TITLE            testin precursor - rat
ALTERNATE_NAMES  testin 1; testis II
ORGANISM         #formal_name Rattus norvegicus #common
DATE             02-Aug-1996 #sequence_revision 02-Aug
                   18-Jun-1999
ACCESSIONS       I52525; A41999; B41999; B01250
REFERENCE        I52525
   #authors      Grima, J.; Zhu, L.J.; Zong, S.D.; Catt
                   C.W.; Cheng, C.Y.
   #journal      Biol. Reprod. (1996) 52:340-355
   #title        Rat testin is a newly identified compo
                   complexes in various tissues whose p
                   expressed in the testis and ovary.
   #cross-references MUID:96227
   #accession    I52525
      ##status         preliminary; translated from GB/E
      ##molecule_type mRNA
      ##residues       1-333 ##label RES
      ##cross-references EMBL:U05058; NID:g577429; PIDN
                        PID:g577430
REFERENCE        A92740
   #authors      Cheng, C.Y.; Grima, J.; Stahler, M.S.;
                   Bardin, C.W.
   #journal      J. Biol. Chem. (1989) 264:21386-21393
   #title        Testins are structurally related Serto
                   secretion is tightly coupled to the
   #cross-references MUID:90078247
   #accession    A34199
      ##molecule_type protein
      ##residues       20-349 ##label CHE
      ##note           this secreted form was designated
   #accession    B34199
      ##molecule_type protein
      ##residues       'X',1e-19,'X',21-28,'X',30-41,'X'
      ##note           this secreted form was designated
COMMENT          This junctional complex protein, although cl
                   cathepsin L, is unlikely to be active as a
CLASSIFICATION   #superfamily papain
KEYWORDS         glycoprotein; monomer; testis
FEATURE
   1-16                #domain signal sequence #status p
   17-333              #product testin II #status predict
   20-333              #product testin II #status predict
   138,276,300         #region detective catalytic triad
   135-178,169-211,
```

```
   269-322              #disulfide_bonds #status predicted\
   173                  #binding_site carbohydrate (Asn) (covalent) #status
                          predicted
SUMMARY            #length 333  #molecular-weight 38033  #checksum 2710

  Query Match             19.1%;  Score 76;  DB 1;  Length 333;
  Best Local Similarity   36.1%;  Pred. No. 2.88e+00;
  Matches    13;  Conservative    8;  Mismatches  15;  Indels   0;  Gaps   0;

Db     200 ESYPYRGQGRECRYHAENSAANVRDFVQIPGSEEAL 235
           |:|  |  :| | ||: :||  : :  || |: : |
Qy       4 PGYLSRNSSRVCNYHGCSSATGILPACQIYCNSNQL 39


RESULT   15
ENTRY              C64725      #type complete
  TITLE            fixC protein - Escherichia coli
CONTAINS           probable quinone reductase (EC 1.6.5.-)
ORGANISM           #formal_name Escherichia coli
DATE               12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
                     20-Sep-1999
ACCESSIONS         C64725; S40564
REFERENCE          A64720
   #authors        Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
                     Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
                     Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
                     Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
                     Y.
   #journal        Science (1997) 277:1453-1462
   #title          The complete genome sequence of Escherichia coli K-12.
   #cross-references MUID:97426617
   #accession      C64725
      ##status          nucleic acid sequence not shown; translation not shown
      ##molecule_type DNA
      ##residues        1-428 ##label BLAT
      ##cross-references GB:AE000114; GB:U00096; NID:g1786217;
                          PIDN:AAC73154.1; PID:g1786227; UWGP:b0043
      ##experimental_source strain K-12, substrain MG1655
REFERENCE          S40531
   #authors        Yura, T.; Mori, H.; Nagai, H.; Nagata, T.; Ishihama, A.;
                     Fujita, N.; Isono, K.; Mizobuchi, K.; Nakata, A.
   #submission     submitted to the EMBL Data Library, December 1992
   #description    Systematic sequencing of the Escherichia coli genome:
                     analysis of the 0-2.4min region.
   #accession      S40564
      ##status          preliminary
      ##molecule_type DNA
      ##residues        1-279,'1M',282-309,'F',311-428 ##label YUR
      ##cross-references EMBL:D10483; NID:g216434; PIDN:BAA01819.1;
                          PID:d1001791; PID:g216468
GENETICS
   #gene           fixC
CLASSIFICATION     #superfamily fixC protein
KEYWORDS           FAD; flavoprotein, oxidoreductase; transmembrane protein
FEATURE
   8-24                 #domain transmembrane #status predicted #label TMM
SUMMARY            #length 428  #molecular-weight 45702  #checksum 2573

  Query Match             19.1%;  Score 76;  DB 2;  Length 428;
  Best Local Similarity   36.4%;  Pred. No. 2.88e+00;
  Matches    16;  Conservative   11;  Mismatches  14;  Indels   3;  Gaps   3;

Db     195 VIEDRFQLQGNQGAACLFAG-SPTDGLMGGGFLYTNENTLSLGL 237
           |||  : |  | :  | : | |:: |:: :| :| | | |||::
Qy       2 VIEC-Y-LSKNSSRVCNYHGCSSATGILFACQIYCNSNQLSLSM 43


Search completed: Fri Aug 25 18:07:42 2000
Job time : 21 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 25 18:08:01 2000; MasPar time 3.72 Seconds
449.571 Million cell updates/sec

Tabular output not generated.

Title: >US-09-257-585-8
Description: (1-54) from US09257585.pep
Perfect Score: 398
Sequence: 1 MVIEGYLSRNSSRVCNYHGC.....NSNQLSLSMDSPVEWRRRRI 54

Scoring table: PAM 150
Gap 11

Searched: 85661 seqs, 40989116 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 36.454, Variance 52.528; scale 0.694

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description	Pred. No.
1	94	23.6	496	1	SRM_MOUSE	TYROSINE-PROTEIN KINAS	4.78e-04
2	83	20.9	1559	1	CLIB_SYNY3	FERREDOXIN DEPENDENT G	5.17e-02
3	80	20.1	1331	1	CYAB_LEIDO	RECEPTOR-TYPE ADENYLAT	1.74e-01
4	79	19.8	212	1	LIP_BACSU	LIPASE PRECURSOR (EC 3	2.59e-01
5	76	19.1	278	1	IE68_VZVD	GENE 63/70 PROTEIN.	8.34e-01
6	76	19.1	333	1	TES1_RAT	TESTIN 1/2 PRECURSOR (	8.34e-01
7	76	19.1	429	1	FIXC_ECOLI	FIXC PROTEIN.	8.34e-01
8	76	19.1	758	1	FCUA_YEREN	FERRICHROME RECEPTOR F	8.34e-01
9	76	19.1	801	1	VP34_SCHPO	PHOSPHATIDYLINOSITOL 3	8.34e-01
10	75	18.8	425	1	GCH2_MYCTU	RIBOFLAVIN BIOSYNTHESI	1.22e+00
11	75	18.8	863	1	YS97_CAEEL	HYPOTHETICAL 97.2 KDA	1.22e+00
12	75	18.8	1458	1	ZZ61_HUMAN	ZINC FINGER PROTEIN 26	1.22e+00
13	74	18.6	808	1	GLND_MYCTU	[PROTEIN-PII] URIDYLYL	1.78e+00
14	73	18.3	257	1	ADH1_CERCA	ALCOHOL DEHYDROGENASE	2.60e+00
15	73	18.3	566	1	YCD6_YEAST	HYPOTHETICAL 64.4 KDA	2.60e+00
16	72	18.1	227	1	FYRC_PSEAE	PROBABLE PYRF-YRE 2F	3.76e+00
17	72	18.1	276	1	CPCE_SYNEL	PHYCOCYANOBILIN LYASE	3.76e+00
18	72	18.1	412	1	MOEA_SALTY	MOLYBDOPTERIN BIOSYNTH	3.76e+00
19	72	18.1	5147	1	FA1_DROME	CADHERIN-RELATED TUMOR	3.76e+00
20	71	17.8	214	1	VIF_SIVMK	VIRION INFECTIVITY FAC	5.42e+00
21	71	17.8	227	1	UNG_MYCTU	URACIL DNA GLYCOSYLASE	5.42e+00
22	71	17.8	305	1	FMT_HELPJ	METHIONYL TRNA FORMYLT	5.42e+00
23	71	17.8	419	1	PA1_SERLI	EXTRACELLULAR PHOSPHOL	5.42e+00
24	71	17.8	515	1	CBIL_PAGMA	CYTOCHROME	5.42e+00
25	71	17.8	559	1	FMGI_MESCR	2,4 BISFDG	5.42e+00
26	70	17.6	119	1	ILMC_BRARE	BETA-2 MICR	7.06e+00
27	70	17.6	204	1	CRAL_CAEEL	CRA_1 FIMBR	7.06e+00
28	70	17.6	329	1	NEPV_ANASP	HOMOCITRATE	7.06e+00
29	70	17.6	389	1	GSTM_PIG	GLYCINE AM	7.06e+00
30	70	17.6	484	1	NUOM_TRYBB	NADH UBIQUI	7.06e+00
31	70	17.6	497	1	YTPK_PHIPN	PROBABLE AM	7.06e+00
32	70	17.6	700	1	KPPI_ASPNG	6-PHOSPHOFR	7.06e+00
33	70	17.6	827	1	RED1_YEAST	RED1 PROTEI	7.06e+00
34	70	17.6	957	1	GVRA_BACSU	EXCINUCLEAS	7.06e+00
35	70	17.6	971	1	VE2_EBDV1	OUTER CAPSI	7.06e+00
36	70	17.6	1951	1	SIN3_RAT	SODIUM CHAN	7.06e+00
37	69	17.3	116	1	B2MG_CYPCA	BETA-2-MICR	9.22e+00
38	69	17.3	357	1	AUX1_ICBAC	AUXIN INDU	9.22e+00
39	69	17.3	439	1	KMOS_RAT	MOS PROTEIN	9.22e+00
40	69	17.3	486	1	CISY_THIFE	CITRATE SYN	9.22e+00
41	69	17.3	491	1	PDK_CAEEL	PROBABLE [P	9.22e+00
42	69	17.3	535	1	GRB7_MOUSE	GROWTH FACT	9.22e+00
43	69	17.3	808	1	VGLH_HSVTU	GLYCOPROTEI	9.22e+00
44	69	17.3	1178	1	VG56_HSV11	HYPOTHETICA	9.22e+00
45	69	17.3	1436	1	WCII_BOVIN	ANTIGEN WC	9.22e+00

ALIGNMENTS

RESULT 1
ID SRM_MOUSE STANDARD; PRT; 496 AA.
AC Q62270; Q62360;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE TYROSINE-PROTEIN KINASE SRM (EC 2.7.1.112) (P1K70.
GN SRMS OR SRM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrat;
OC Mammalia; Eutheria; Rodentia; Sciuromorphi; Muri;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=THYMUS;
RX MEDLINE; 97369678.
RA Kawachi Y., Nakauchi H., Tsuka F.;
RT "Isolation of a cDNA encoding a tyrosine kinase ex
RT skin.";
RL Exp. Dermatol. 21:533-538(1995)
RN [2]
RP SEQUENCE FROM N.A
RC TISSUE=LUNG;
RX MEDLINE; 95021220.
RA Kohmura N., Yagi T., Tomooka Y., Oyanagi M., Kobi
RA Chiba J., Ikawa Y., Aizawa S.;
RT "A novel nonreceptor tyrosine kinase, Srm: c.din
RT disruption.";
RL Mol. Cell. Biol. 14:6915-6925(1994).
CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE
CC PROTEIN TYROSINE PHOSPHATE.
CC ! SIMILARITY: TO OTHER PROTEIN TYROSINE KINASES
CC DOMAIN. BELONGS TO THE SRC SUBFAMILY.
CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC ! SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC ----------
CC This SWISS-PROT entry is copyright. It is produced
CC between the Swiss Institute of Bioinformatics an
CC the European Bioinformatics Institute. There are
CC use by non-profit institutions as long as its
CC modified and this statement is not removed. Usage
CC entities requires a license agreement (See http:/
CC or send an email to license@isb-sib.ch).
CC ----------
DR EMBL; D49427; BAA08404.1;
DR EMBL; D26186; BAA05431.1;
DR HSSP; P11362; 1FG1.
DR MGD; MGI:101865; SRMS.

```
DR   PFAM; PF00017; SH2; 1.
DR   PFAM; PF00018; SH3; 1.
DR   PFAM; PF00069; pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   PRINTS; PR00401; SH2DOMAIN.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS50001; SH2; 1.
DR   PROSITE; PS50002; SH3; 1.
KW   Transferase; Tyrosine-protein kinase; ATP-binding; SH2 domain;
KW   SH3 domain; Phosphorylation.
FT   DOMAIN       55    116       SH3.
FT   DOMAIN      124    216       SH2.
FT   DOMAIN      234    495       PROTEIN KINASE.
FT   NP_BIND     240    248       ATP (BY SIMILARITY).
FT   BINDING     262    262       ATP (BY SIMILARITY).
FT   ACT_SITE    354    354       BY SIMILARITY.
FT   MOD_RES     384    384       PHOSPHORYLATION (AUTO?) (BY SIMILARITY)
FT   CONFLICT     78     78       R => G (IN REF. 2).
FT   CONFLICT    236    238       DRK => FGR (IN REF. 2).
FT   CONFLICT    278    278       N => I (IN REF. 2).
SQ   SEQUENCE   496 AA;  55731 MW;  0BD9DB7DB363F3D1 CRC64;

  Query Match           24.6%;  Score 94;  DB 1;  Length 496;
  Best Local Similarity 36.6%;  Pred. No. 4.78e-04;
  Matches   15;  Conservative   12;  Mismatches  10;  Indels   1;  Gaps   1;

Db    157 IGGYSISVRAQAKVCHYRICMAPSGSLYIQFGQLEFSIDAL 197
          | || || |  ::||:|: | :: | : |: ||:: : : |
Qy      3 IEGY-LS-RNSSRVCNYHGCSSAIG-I-LEAGQIYGNSNQL 39


RESULT    2
ID   GLTB_SYNY3     STANDARD;      PRT;  1550 AA.
AC   P55037;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   FERREDOXIN-DEPENDENT GLUTAMATE SYNTHASE 1 (EC 1.4.7.1) (FD-GOGAT).
GN   GLTB OR SLL1502.
OS   Synechocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95244836.
RA   Navarro F., Chavez S., Candau P., Florencio F.J.;
RT   "Existence of two ferredoxin glutamate synthases in the
RT   cyanobacterium Synechocystis sp. PCC 6803. Isolation and insertional
RT   inactivation of gltB and gltS genes.";
RL   Plant Mol. Biol. 27:753-767(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97061201.
RA   Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA   Miyajima N., Hirosawa M., Sugiura M., Sasamoto S., Kimura T.,
RA   Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,
RA   Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
RA   Yamada M., Yasuda M., Tabata S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT   entire genome and assignment of potential protein-coding regions.";
RL   DNA Res. 3:109-136(1996).
CC   -!- CATALYTIC ACTIVITY: 2 L-GLUTAMATE + 2 OXIDIZED FERREDOXIN =
CC       L-GLUTAMINE + 2-OXOGLUTARATE + 2 REDUCED FERREDOXIN.
CC   -!- COFACTOR: BINDS A 3FE-4S CLUSTER, FAD AND FMN.
CC   -!- PATHWAY: GLUTAMINE SYNTHETASE/GOGAT PATHWAY WHICH IS INVOLVED
CC       IN THE ASSIMILATION OF AMMONIA.
CC   -!- SIMILARITY: TO OTHER GLUTAMATE SYNTHASES.
CC   -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-21 IS THE INITIATOR.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X80485; CAA56652.1; -.
DR   EMBL; D90902; BAA17018.1; -.
DR   HSSP; P02392; 1CTF.
DR   PFAM; PF01493; DUF14; 1.
DR   PFAM; PF01645; Glu_synthase; 1.
KW   Oxidoreductase; Iron-sulfur; 3Fe-4S; Flavoprotein; FAD; FMN;
KW   Glutamate biosynthesis.
FT   DOMAIN       43    393       GLUTAMINE AMIDOTRANSFERASE (POTENTIAL).
FT   NP_BIND    1097   1154       FMN (BY SIMILARITY).
FT   METAL      1150   1150       IRON-SULFUR (3FE-4S) (BY SIMILARITY).
FT   METAL      1156   1156       IRON-SULFUR (3FE-4S) (BY SIMILARITY).
FT   METAL      1161   1161       IRON-SULFUR (3FE-4S) (BY SIMILARITY).
SQ   SEQUENCE   1550 AA;  169071 MW;  97AA64D832296943 CRC64;

  Query Match           20.9%;  Score 83;  DB 1;  Length 1550;
  Best Local Similarity 26.4%;  Pred. No. 5.17e-02;
  Matches   14;  Conservative   15;  Mismatches  21;  Indels   3;  Gaps   3;

Db    777 VIDEYFCRTSSPTQGSDLGVIAQEAILRHQHAFAPPPQDLHTTDVQGEYQWRK 829
          || | | ||| :    | : :|| : :: : :| :| : : :||:
Qy      2 VIEGYLSRNSSRVCNYH GCSSAIGILEAGQIYG NSNQL-SLSMDSPVEWRR 51


RESULT    3
ID   CYAB_LEIDO     STANDARD;      PRT;  1331 AA.
AC   Q25263;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   RECEPTOR-TYPE ADENYLATE CYCLASE B (EC 4.6.1.1) (ATP PYROPHOSPHATE-
DE   LYASE) (ADENYLYL CYCLASE).
GN   RAC B.
OS   Leishmania donovani.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1S SUDANESE;
RX   MEDLINE; 95340554.
RA   Sanchez M.A., Zeuli D., Kiamo E.M., Kavanaugh M.P., Landfear S.M.;
RT   "A family of putative receptor adenylate cyclases from Leishmania
RT   donovani.";
RL   J. Biol. Chem. 270:17551-17558(1995).
CC   -!- FUNCTION: COULD ACT AS A RECEPTOR FOR A UNKNOWN LIGAND.
CC    !- CATALYTIC ACTIVITY: ATP = 3',5'-CYCLIC AMP + PYROPHOSPHATE.
CC   -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED IN THE INSECT STAGE (PROMASTIGOTE)
CC       BUT NOT IN THE MAMMALIAN HOST STAGE OF THE PARASITE LIFE CYCLE.
CC    !- SIMILARITY: BELONGS TO ADENYLYL CYCLASE CLASS-3 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U17043; AAA74999.1; -.
DR   PFAM; PF00211; guanylate_cyc; 1.
KW   Lyase; cAMP synthesis; Transmembrane; Receptor; Glycoprotein.
FT   DOMAIN        1     33       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     34     54       POTENTIAL.
FT   DOMAIN       55    898       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    899    919       POTENTIAL.
FT   DOMAIN      920   1331       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD    255    255       POTENTIAL.
FT   CARBOHYD    429    429       POTENTIAL.
```

```
FT   CARBOHYD     558     558       POTENTIAL.
FT   CARBOHYD     574     574       POTENTIAL.
FT   CARBOHYD     657     657       POTENTIAL.
SQ   SEQUENCE   1331 AA;  144162 MW;  5F03AA22DC1B5072 CRC64

  Query Match           29.1%;  Score 80;  DB 1,  Length 1331;
  Best Local Similarity  25.5%;  Pred. No. 1.74e-01;
  Matches   13;  Conservative   14;  Mismatches  22;  Indels   2;  Gaps   2;

Db    837 GCTTTPTDRFCVYHNYGAQGIVMLSVQRMLDPIVPQLSSPMTPTMDYRPRQ 887
          |  :  : | | ||  :: :| :|    ::   :  ||| :| : :: | |
Qy      5 GYLSPNSSPVCNYHGCSS-AIGILEAVQIYCNS NQLSLSMDSPVEWRRPP 53


RESULT    4
ID   LIP_BACSU      STANDARD;      PRT;   212 AA.
AC   P37957; O34644;
DT   01-OCT-1994 (Rel. 30, Created)
DT   15-FEB-2000 (Rel. 39, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   LIPASE PRECURSOR (EC 3.1.1.3) (TRIACYLGLYCEROL LIPASE).
GN   LIPA OR LIP.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE; 92329538.
RA   Dartois V., Baulard A., Schanck K., Colson C.,
RT   "Cloning, nucleotide sequence and expression in Escherichia coli of a
RT   lipase gene from Bacillus subtilis 168.";
RL   Biochim. Biophys. Acta 1131:253-260(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   Kumano M., Tamakoshi A., Yamane K.;
RT   "A 32 kb nucleotide sequence from the region of the lincomycin
RT   resistance gene (22-25 degree) of the Bacillus subtilis chromosome and
RT   identification of the site of the lin-2 mutation.";
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: TRIACYLGLYCEROL + H(2)O = DIACYLGLYCEROL +
CC       A FATTY ACID ANION.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   ---------- ------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ------------------------------------------------------------
DR   EMBL; M74010; AAA22574.1; -.
DR   EMBL; AB000617; BAA22231.1; -.
DR   EMBL; Z99105; CAB12064.1; -.
DR   PIR; S23934; S23934.
DR   SUBTILIST; BG10679; LIPA.
DR   PFAM; PF01674; Lipase_2; 1.
KW   Hydrolase; Lipid degradation; Signal.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25    212       LIPASE.
FT   ACT_SITE     108    108       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CONFLICT     133    133       L -> V (IN REF. 1).
SQ   SEQUENCE   212 AA;  22791 MW;  B8A70E027461188F CRC64;

  Query Match           19.8%;  Score 79;  DB 1;  Length 212;
  Best Local Similarity  36.4%;  Pred. No. 2.59e-01;
  Matches   12;  Conservative   10;  Mismatches   9;  Indels   2;  Gaps   2;

Db    165 MIVMNYLSRELDGARNVQTHCVGH-IGLLYSSQV 196
          |::  |||| :::|    || :  |||| :::|:
Qy      1 MVIEGYLSR-NSSRVCNYHGCSSAIGILEAGQI 32


RESULT    5
ID   IE68_VZVD      STANDARD;      PRT;   278 AA.
AC   P09255;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   GENE 63/70 PROTEIN.
GN   63 OR 70.
OS   Varicella-zoster virus (strain Dumas) (VZV).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesvirid
OC   Alphaherpesvirinae; Varicellovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 86306657.
RA   Davison A.J., Scott J.E.;
RT   "The complete DNA sequence of varicella-zoster vi
RL   J. Gen. Virol. 67:1759-1816(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 85107111.
RA   Davison A.J., Scott J.E.;
RT   "DNA sequence of the major inverted repeat in the
RT   virus genome.";
RL   J. Gen. Virol. 66:207-220(1985).
CC   -!- SIMILARITY: BELONGS TO A FAMILY THAT GROUP
CC       HSV-2 IE-68 (US1), EHV-1 65, EHV-4 (ORF4), PRV
CC   ---------- ----------
CC   This SWISS-PROT entry is copyright. It is produced
CC   between the Swiss Institute of Bioinformatics an
CC   the European Bioinformatics Institute. There are
CC   use by non-profit institutions as long as its
CC   modified and this statement is not removed. Usage
CC   entities requires a license agreement (See http://
CC   or send an email to license@isb-sib.ch).
CC   ---------- ----------
DR   EMBL; X04370; CAA27944.1; -.
DR   EMBL; X04370; CAA27945.1; -.
DR   EMBL; X02142; CAA26045.1; -.
DR   PIR; B27345; EDBE63.
KW   Early protein.
SQ   SEQUENCE   278 AA;  30495 MW;  2CFA03C199F9A485 CR

  Query Match           19.1%;  Score 76;  DB 1;  Lengt
  Best Local Similarity  30.8%;  Pred. No. 8.44e-01;
  Matches    8;  Conservative   10;  Mismatches   7;

Db    229 ETAE YTAAKALTALEKEVWKEER 263
          |::: |  :: |:  :: : : :|
Qy     28 PAQ'Y'NQNGELDM'YSW 48


RESULT    6
ID   TES1_RAT       STANDARD;      PRT;   833 AA.
AC   P15242; P15243;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   TESTIN 1/2 PRECURSOR (PMP 22/PMP-23).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RX   MEDLINE; 95226598.
RA   Grima J., Zhu L., Zoud S.D., Waterall J.P., Bardi
RA   Cheng C.Y.;
RT   "Rat testin is a newly identified component of the
RT   complexes in various tissues whose mRNA is predom
RT   the testis and ovary.";
RL   Biol. Reprod. 52:340-[illegible]
```

```
RN   [2]
RP   SEQUENCE OF 18-49.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=SERTOLI CELLS;
RX   MEDLINE; 90078247.
RA   Cheng C.Y., Grima J., Stahler M.S., Lookshin R.A.;
RT   "Testins are structurally related Sertoli cell proteins whose
RT   secretion is tightly coupled to the presence of germ cells.";
RL   J. Biol. Chem. 264:21386-21393(1989).
RN   [3]
RP   LACK OF PROTEASE ACTIVITY.
RX   MEDLINE; 94191698.
RA   Cheng C.Y., Morris I., Bardin C.W.;
RT   "Testins are structurally related to the mouse cysteine proteinase
RT   precursor but devoid of any protease/anti-protease activity.";
RL   Biochem. Biophys. Res. Commun. 191:224-231(1993).
CC   -!- FUNCTION: NOT KNOWN.
CC   -!- SUBCELLULAR LOCATION: SERTOLI CELLS SECRETORY PROTEIN.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1; ALSO KNOWN AS THE
CC       PAPAIN FAMILY OF THIOL PROTEASES.
CC   -!- CAUTION: SHOULD NOT BE CONFUSED WITH MOUSE TESTIN WHICH IS A
CC       LIM DOMAIN PROTEIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U16858; AAC52162.1; -.
DR   PIR; A34199; A34199.
DR   PIR; B34199; B34199.
DR   HSSP; P07711; 1CJL.
DR   PFAM; PF00112; Peptidase_C1; 1.
DR   PRINTS; PR00705; PAPAIN.
DR   PROSITE; PS00139; THIOL_PROTEASE_CYS; PARTIAL_MSG.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
DR   PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
KW   Testis; Signal.
FT   SIGNAL       1     17
FT   CHAIN       18    333       TESTIN 2.
FT   CHAIN       20    333       TESTIN 1.
FT   SITE       138    138       ANCESTRAL ACTIVE SITE.
FT   ACT_SITE   276    276       BY SIMILARITY.
FT   ACT_SITE   300    300       BY SIMILARITY.
FT   DISULFID   135    178       BY SIMILARITY
FT   DISULFID   169    211       BY SIMILARITY.
FT   DISULFID   269    322       BY SIMILARITY.
SQ   SEQUENCE   333 AA;  38033 MW;  B4426BD8B2B66A2E CRC64;

  Query Match          19.1%;  Score 76;  DB 1;  Length 333;
  Best Local Similarity  36.1%;  Pred. No. 8.34e-01;
  Matches   13;  Conservative    8;  Mismatches  15;  Indels   0;  Gaps   0;

Db     200 ESYPYRGQGRECRYHAENSAANVRDFVQIPCSEEAL 239
           |:|  |  :| | ||: :||  : :  || |: : |
Qy       4 EGYISRNSSRVCNYHGCSSAIGILEAGQIYGNSNQL 39


RESULT    7
ID   FIXC_ECOLI     STANDARD;      PRT;   428 AA.
AC   P31575; P75626;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   FIXC PROTEIN.
GN   FIXC.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE; 92334977.
RA   Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,
RA   Isono K., Mizobuchi K., Nakata A.;
RT   "Systematic sequencing of the Escherichia coli genome: analysis of
RT   the 0-2.4 min region.";
RL   Nucleic Acids Res. 20:3305-3308(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN K12 / MG1655;
RX   MEDLINE; 97426617.
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [3]
RP   SEQUENCE OF 1-188 FROM N.A.
RC   STRAIN=044 K74;
RX   MEDLINE; 96066354.
RA   Eichler K., Buchet A., Bourgis F., Kleber H.-P.,
RA   Mandrand-Berthelot M.-A.;
RT   "The fix Escherichia coli region contains four genes related to
RT   carnitine metabolism.";
RL   J. Basic Microbiol. 35:217-227(1995).
CC   -!- FUNCTION: PROBABLY ACCEPTS ELECTRONS FROM FIXA/FIXB AND REDUCES A
CC       QUINONE.
CC   -!- COFACTOR: FAD (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE ETF-QO / FIXC FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D10483; CAB20038.1; -.
DR   EMBL; AE000114; AAC73154.1; -.
DR   EMBL; X71977; CAA50799.1; -.
DR   PIR; S40564; S40564.
DR   HSSP; P15245; 1FOH.
DR   ECOGENE; EG11564; FIXC.
DR   PFAM; PF01494; FAD_binding_3; 1.
KW   Oxidoreductase; FAD; Flavoprotein.
FT   NP_BIND     12     28       FAD OR NAD(P) (ADP PART) (POTENTIAL).
FT   CONFLICT     2      3       SE -> T (IN REF. 3).
FT   CONFLICT    64     64       D -> E (IN REF. 1).
FT   CONFLICT   280    281       VV -> IM (IN REF. 1).
FT   CONFLICT   310    310       L -> F (IN REF. 1).
SQ   SEQUENCE   428 AA;  45702 MW;  4D089EE9A3007FBC CRC64;

  Query Match          19.1%;  Score 76;  DB 1;  Length 428;
  Best Local Similarity  36.4%;  Pred. No. 8.34e-01;
  Matches   16;  Conservative   11;  Mismatches  14;  Indels   3;  Gaps   3;

Db     195 VIEDRFQLQGNQGAACLFAG-SPTDGLMGGGFLYTNENTLSLGL 237
           |||  : |  | :  | : | |:: :: :| :| | | |||::
Qy       2 VIEG-Y-LSRNSSRVCNYHGCSSAIGILEAGQIYGNSNQLSLSM 43


RESULT    8
ID   FCUA_YEREN     STANDARD;      PRT;   758 AA.
AC   Q05202;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   FERRICHROME RECEPTOR FCUA PRECURSOR.
GN   FCUA.
OS   Yersinia enterocolitica.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
```

```
OC   Yersinia.
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-5 AND 37-46.
RC   STRAIN=SEROTYPE O:8;
RX   MEDLINE; 93211281.
RA   Koebnik R., Hantke K., Braun V.;
RT   "The TonB-dependent ferrichrome receptor FcuA of Yersinia
RT   enterocolitica: evidence against a strict co-evolution of receptor
RT   structure and substrate specificity.";
RL   Mol. Microbiol. 7:383-393(1993).
CC   -!- FUNCTION: RECEPTOR FOR THE HYDROXAMATE SIDEROPHORE, FERRICHROME.
CC       BINDS ALSO TO MOST OTHER FERRICHROME DERIVATIVES EXCEPT ENANTIO
CC       FERRICHROME AND FERRIC RHODOTORULATE.
CC   -!- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC   -!- INDUCTION: BY IRON-STARVATION CONDITIONS. IRON REGULATION
CC       MEDIATED THROUGH THE FUR PROTEIN.
CC   -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC   ---------- ----------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ----------
DR   EMBL; X67331; CAA47746.1; -.
DR   PFAM; PF00593; TonB_boxC; 1.
DR   PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
DR   PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW   Signal; Receptor; Transmembrane; Outer membrane; Iron transport;
KW   Transport; TonB box.
FT   SIGNAL       1     36
FT   CHAIN       37    758       FERRICHROME RECEPTOR FCUA.
FT   SIMILAR     66     73       TONB BOX.
FT   SIMILAR    741    758       TONB C-TERMINAL BOX.
SQ   SEQUENCE   758 AA;  81749 MW;  945675B0DEF39F7C CRC64;

  Query Match             19.1%;  Score 76;  DB 1;  Length 758;
  Best Local Similarity   31.4%;  Pred. No. 8.34e-01;
  Matches   11;  Conservative   13;  Mismatches   8;  Indels   3;  Gaps   3;

Db    224 PRVTVDY GSASSVGGALDVGNRYGDGAQFGVRVN 257
          ||  :| | :| :|  :| |: |: || :|::: ::
Qy     12 SRV-CNYHGCSSAIG-ILEAGQIYGNSNQLSLSMD 44


RESULT    9
ID   VPS4_SCHPO     STANDARD;      PRT;   801 AA
AC   P50520;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   PHOSPHATIDYLINOSITOL 3-KINASE VPS34 (EC 2.7.1.137) (PI3-KINASE)
DE   (PTDINS-3-KINASE) (PI3K) (VACUOLAR SORTING PROTEIN 34).
GN   VPS34.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
OC   Schizosaccharomycetaceae; Schizosaccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96360252.
RA   Takegawa K., Dewald D.B., Emr S.E.;
RT   "Schizosaccharomyces pombe Vps34p, a phosphatidylinositol-specific PI
RT   3-kinase essential for normal cell growth and vacuole morphology.";
RL   J. Cell Sci. 108:3745-3756(1995).
CC   -!- FUNCTION: PHOSPHATIDYLINOSITOL 3-KINASE HOMOLOG REQUIRED FOR
CC       VACUOLAR SORTING AND SEGREGATION.
CC   -!- CATALYTIC ACTIVITY: ATP + 1-PHOSPHATIDYL-1D-MYO-INOSITOL = ADP +
CC       1-PHOSPHATIDYL-1D-MYO-INOSITOL 3-PHOSPHATE.
CC   -!- SIMILARITY: BELONGS TO THE PI3/PI4-KINASES FAMILY.
CC   ---------- ----------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ----------
DR   EMBL; U32593; AAC49133.1; -.
DR   PFAM; PF00792; PI3K_C; 1.
DR   PFAM; PF00613; PI3Ka; 1.
DR   PFAM; PF00454; PI3_PI4_kinase; 1.
DR   PROSITE; PS00915; PI3_4_KINASE_1; 1.
DR   PROSITE; PS00916; PI3_4_KINASE_2; 1.
KW   Transferase; Kinase.
FT   DOMAIN      541    710       PI3K/PI4K.
SQ   SEQUENCE   801 AA;  [illegible] MW;  1FF491E74BAFB4F3 CR

  Query Match             [illegible]%;  Score 76;  DB 1;  Length
  Best Local Similarity   41.7%;  Pred. No. 8.34e-01;
  Matches   10;  Conservative    9;  Mismatches   [illegible]

Db    224 RVTYGSFDLELNLDSAEDLKHEK 247
          ::| :| :| |::||: :: :
Qy     33 SIPQNFNLISLSMGSVEWESEK 54


RESULT   10
ID   SCH2_MYCTO     STANDARD;      PRT:   425 AA
AC   P71684;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   RIBOFLAVIN BIOSYNTHESIS PROTEIN RIBA [INCLUDES: GT[illegible]
DE   (EC 3.5.4.25); 3,4-DIHYDROXY-2 BUTANONE 4-PHOSPHAT[illegible]
DE   SYNTHASE)].
GN   RIBA OR RV1415 OR MTCY20B4.33
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteri[illegible]
OC   Actinomycetales; Corynebacterineae; Mycobacteriace[illegible]
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE; 98295987.
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Chu[illegible]
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III,
RA   Badcock K., Basham D., Brown D., Chillingworth T.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Ham[illegible]
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule[illegible]
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squa[illegible]
RA   Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
CC   -!- CATALYTIC ACTIVITY: GTP + 3 H(2)O = FORMATE + [illegible]
CC       HYDROXY-4 (5-PHOSPHORIBOSYLAMINO)PYRIMIDINE + [illegible]
CC   -!- PATHWAY: RIBOFLAVIN BIOSYNTHESIS.
CC   -!- SIMILARITY: TO OTHER BACTERIAL GTP CYCLOHYDROLA[illegible]
CC       TO E.COLI RIBB AND V.HARVEYI LUXH.
CC   ---------- ----------
CC   This SWISS-PROT entry is copyright. It is produced [illegible]
CC   between the Swiss Institute of Bioinformatics and [illegible]
CC   the European Bioinformatics Institute. There are [illegible]
CC   use by non-profit institutions as long as its [illegible]
CC   modified and this statement is not removed. Usage [illegible]
CC   entities requires a license agreement (See http://[illegible]
CC   or send an email to license@isb-sib.ch).
CC   ---------- ----------
DR   EMBL; Z80108; CAB02199.1; -.
DR   TUBERCULIST; RV1415; -.
DR   PFAM; PF00926; DHBP_synthase; 1.
DR   PFAM; PF00925; GTP_cyclohydro2; 1.
KW   Multifunctional enzyme; Riboflavin biosynthesis; Hy[illegible]
```

```
SQ   SEQUENCE   425 AA;  46017 MW;  FA7F748688145915 CRC64;

  Query Match           18.8%;  Score 75;  DB 1;  Length 425;
  Best Local Similarity  25.0%;  Pred. No. 1.22e+00;
  Matches    6;  Conservative    12;  Mismatches    5;  Indels    1;  Gaps    1;

Db    182 RVFADEHGLALITIADLIEWRRKH 205
          ::::: : |:| ::   :|||||::
Qy     31 QIYGNSNQLSL-SMDSPVEWRRRR 53


RESULT   11
ID   YS97_CAEEL     STANDARD;      PRT;   863 AA.
AC   Q09966;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   HYPOTHETICAL 97.2 KDA PROTEIN B0244.7 IN CHROMOSOME III.
GN   B0244.7.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Favello T., Waterston R.;
RL   Submitted (JUN 1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO A FAMILY THAT CURRENTLY CONSISTS OF
CC       B0244.4, B0244.5, B0244.6 AND B0244.7.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U28971; AAA68378.1; -.
DR   WORMPEP; B0244.7; CE02416.
DR   PFAM; PF00001; 7tm_1; 1.
KW   Hypothetical protein; Transmembrane.
FT   DOMAIN      596    599       POLY-ILE.
FT   TRANSMEM     49     69       POTENTIAL.
FT   TRANSMEM    107    127       POTENTIAL.
FT   TRANSMEM    131    151       POTENTIAL.
FT   TRANSMEM    166    186       POTENTIAL.
FT   TRANSMEM    206    226       POTENTIAL.
FT   TRANSMEM    262    282       POTENTIAL.
FT   TRANSMEM    313    333       POTENTIAL.
FT   TRANSMEM    352    372       POTENTIAL.
FT   TRANSMEM    393    413       POTENTIAL.
FT   TRANSMEM    536    556       POTENTIAL.
FT   TRANSMEM    579    599       POTENTIAL.
FT   TRANSMEM    612    632       POTENTIAL.
FT   TRANSMEM    660    680       POTENTIAL.
FT   TRANSMEM    698    718       POTENTIAL.
FT   TRANSMEM    749    769       POTENTIAL.
FT   TRANSMEM    787    807       POTENTIAL.
SQ   SEQUENCE   863 AA;  97228 MW;  B56BB2B6432768BD CRC64;

  Query Match           18.8%;  Score 75;  DB 1;  Length 863;
  Best Local Similarity  29.6%;  Pred. No. 1.22e+00;
  Matches    8;  Conservative    12;  Mismatches    7;  Indels    0;  Gaps    0;

Db    185 IVEAAFIGRNSPFSCDYGNCAGPVLIF 211
          :|  ::::|||:  |:|  |:::: |:
Qy      1 MVIFGYISRNSSRVCNYHGCSSAIGIL 27


RESULT   12
ID   Z261_HUMAN     STANDARD;      PRT;  1358 AA.
AC   Q14202;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   ZINC FINGER PROTEIN 261 (DXS6673E PROTEIN).
GN   ZNF261.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=FETAL BRAIN;
RX   MEDLINE; 96414295.
RA   van der Maarel S.M., Scholten I.H.G.M., Huber I., Phillippe C.,
RA   Suijkerbuijk R.F., Gilgenkrantz S., Kere J., Cremers F.P.M.,
RA   Ropers H.-H.;
RT   "Cloning and characterization of DXS6673E, a candidate gene for
RT   X-linked mental retardation in Xq13.1.";
RL   Hum. Mol. Genet. 5:887-897(1996).
CC   -!- TISSUE SPECIFICITY: MOST ABUNDANT IN BRAIN, MODERATE IN MUSCLE AND
CC       HEART, LOW IN OTHER TISSUES EXCEPT PLACENTA.
CC   -!- DISEASE: INVOLVED IN A T(X;13)(Q13.1;?) CHROMOSOMAL TRANSLOCATION;
CC       POSSIBLE CANDIDATE FOR X LINKED MENTAL RETARDATION IN XQ13.1.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X95808; CAA65075.1; -.
DR   MIM; 300061; -.
KW   Chromosomal translocation; Transmembrane.
FT   TRANSMEM    845    866       POTENTIAL.
SQ   SEQUENCE   1358 AA;  151863 MW;  [illegible] CRC64;

  Query Match           18.8%;  Score 75;  DB 1;  Length 1358;
  Best Local Similarity  31.8%;  Pred. No. 1.22e+00;
  Matches   14;  Conservative   13;  Mismatches   16;  Indels    1;  Gaps    1;

Db    926 MIAEAEELDKASSDLCDLVSNQSAEGLLEDCDLPGPARDDVLAM 969
          |: |:  | : || :|:  :  || |:|| :::| : :  |:|
Qy      1 MVIECY-LSRNSSRVCNYHGCSSAIGILEACQIYGNSNQLSLSM 43


RESULT   13
ID   GLND_MYCTU     STANDARD;      PRT;   808 AA.
AC   Q10961;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   [PROTEIN-PII] URIDYLYLTRANSFERASE (EC 2.7.7.59) (PII URIDYLYL
DE   TRANSFERASE) (URIDYLYL REMOVING ENZYME) (UTASE).
GN   GLND OR RV2918C OR MTCY338.07C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE; 98295987.
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA   Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
```

```
RL   Nature 394:537-544(1998).
CC   -!- FUNCTION: MODIFIES, BY URIDYLYLATION OR DEURIDYLYLATION THE PII
CC       (GLNB) REGULATORY PROTEIN (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: UTP + [PROTEIN-PII] = PYROPHOSPHATE +
CC       URIDYLYL-[PROTEIN-PII].
CC   -!- SIMILARITY: BELONGS TO THE GLND FAMILY.
CC   --------  --------  ----------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------------
DR   EMBL; Z74697; CAA98971.1; -.
DR   TUBERCULIST; RV2918C; -.
DR   PFAM; PF01966; HD; 1.
KW   Transferase; Nucleotidyltransferase.
SQ   SEQUENCE   808 AA;  86437 MW;  9B1F14D01C29F0B0 CRC64;

  Query Match           18.6%;  Score 74;  DB 1;  Length 808;
  Best Local Similarity  40.6%;  Pred. No. 1.78e+00;
  Matches   13;  Conservative   11;  Mismatches   6;  Indels   2;  Gaps   2;

Db    130 AALGMLEARHIAGD-QQLSFALIDGV-RRQWR 159
          :|:|:||| :| |:  ||:::   | :|:::|
Qy     22 SAIGILEAGQIYGNSNQLSLSMDSPVEWRRRR 53


RESULT   14
ID   ADH1_CERCA     STANDARD;      PRT;   257 AA.
AC   P48814;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   ALCOHOL DEHYDROGENASE 1 (EC 1.1.1.1).
GN   ADH1.
OS   Ceratitis capitata (Mediterranean fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Tephritoidea; Tephritidae; Ceratitis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BENAKEION;
RA   Savakis C.;
RL   Submitted (FEB-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: ALCOHOL + NAD(+) = ALDEHYDE OR KETONE + NADH.
CC   -!- SUBUNIT: HOMODIMER.
CC   -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC       (SDR) FAMILY.
CC   ----------  --------  ----------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------------
DR   EMBL; Z30194; CAA82926.1; -.
DR   PFAM; PF00106; adh_short; 1.
DR   PFAM; PF00663; adh_short_C; 1.
DR   PRINTS; PR00080; ALCDHDRGNASE.
DR   PRINTS; PR01167; INSADHFAMILY.
DR   PRINTS; PR01168; ALCDHDRGNASE.
DR   PRINTS; PR01169; CERATITISADH.
DR   PROSITE; PS00061; ADH_SHORT; 1.
KW   Oxidoreductase; NAD.
FT   NP_BIND      9     33       NAD (BY SIMILARITY).
FT   ACT_SITE    150    150      BY SIMILARITY.
SQ   SEQUENCE   257 AA;  27922 MW;  043F39AC1E09CACE CRC64;

  Query Match           18.3%;  Score 73;  DB 1;  Len
  Best Local Similarity  40.5%;  Pred. No. 2.60e+00;
  Matches   15;  Conservative    7;  Mismatches   2;

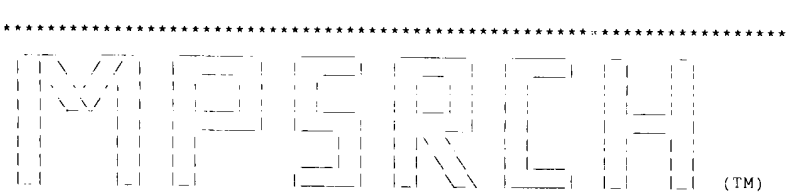
Db    219 VCGAHL-AQAIELMGNCAIY [illegible]
          ||  |  : || ::: | ||  |:| :|:: | | |
Qy     14 VCNYHGCSSAIGILEAGQIY [illegible]

RESULT   15
ID   YCD6_YEAST     STANDARD;      PRT;   566 AA.
AC   P25370;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   HYPOTHETICAL 64.9 KDA PROTEIN IN SRO9-SIE50 INTER
GN   YCL036W OR YCL36W OR YCL187.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Sa
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91377317.
RA   Rad M.R., Indzeykiriben F., Xi J., Kleinhans U.,
RT   "The complete sequence of a 11,453 bp fragment fr
RT   III encompasses four new open reading frames.";
RL   Yeast 7:533-538(1991).
CC   --------------------  ---  -  ----------
CC   This SWISS-PROT entry is copyright. It is produc
CC   between the Swiss Institute of Bioinformatics a
CC   the European Bioinformatics Institute. There are
CC   use by non-profit institutions as long as its
CC   modified and this statement is not removed. Usage
CC   entities requires a license agreement (See http:/
CC   or send an email to license@isb-sib.ch).
CC   --------------------  ---  -  ----------
DR   EMBL; X59720; CAA42386.1; -.
DR   PIR; S19364; S19364.
DR   PIR; S17477; S17477.
KW   Hypothetical protein.
SQ   SEQUENCE   566 AA;  [illegible] MW;  [illegible] CRC64;

  Query Match           18.3%;  Score 73;  DB 1;  Leng
  Best Local Similarity  42.9%;  Pred. No. 2.60e+00;
  Matches    9;  Conservative    4;  Mismatches   2;

Db    [illegible]
          ::|| :|  :|  :|:|:| :|
Qy      1 MVIEGYLS RNSSRVCNYIDT 20


Search completed: Fri Aug 25 [illegible] 2000
Job time : 12 secs.
```



Release 4.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 25 18:08:31 2000; MasPar time 8.67 Seconds
431.818 Million cell updates/sec

Tabular output not generated.

Title: US-09-257-585-8
Description: (1-54) from US09257585.pep
Perfect Score: 398
Sequence: 1 MVIEGYLSRNSSRVCNYHGC.....NSNQLSLSMDSPVFWRRRRI 54

Scoring table: PAM 150
Gap 11

Searched: 225879 seqs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl12
1:sp_archea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 35.558; Variance 54.499; scale 0.652

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	90	22.6	314	3	O74544	YEAST GTR2 HOMOLOG, NO	9.93e-03
2	86	21.6	535	14	Q66188	(STRAIN D1466) PEPLOME	5.05e-02
3	86	21.6	758	14	Q97628	CAC (FRAGMENT).	5.05e-02
4	86	21.6	858	14	Q12002	ENVELOPE GLYCOPROTEIN	5.05e-02
5	86	21.6	874	14	Q12003	ENVELOPE GLYCOPROTEIN	5.05e-02
6	85	21.4	171	10	Q40966	ALBUMIN 2 (FRAGMENT).	7.54e-02
7	85	21.4	862	14	Q12004	ENVELOPE GLYCOPROTEIN	7.54e-02
8	85	21.4	870	14	Q12011	ENVELOPE GLYCOPROTEIN	7.54e-02
9	84	21.1	862	14	Q12010	ENVELOPE GLYCOPROTEIN	1.12e-01
10	83	20.9	854	14	Q12008	ENVELOPE GLYCOPROTEIN	1.66e-01
11	83	20.9	866	14	Q12006	ENVELOPE GLYCOPROTEIN	1.66e-01
12	83	20.9	868	14	Q12009	ENVELOPE GLYCOPROTEIN	1.66e-01
13	83	20.9	2081	5	Q9XXY8	L302.11 PROTEIN.	1.66e-01
14	82	20.6	683	1	Q9YBJ9	683AA LONG HYPOTHETICA	2.46e-01
15	82	20.6	866	14	Q12005	ENVELOPE GLYCOPROTEIN	2.46e-01
16	81	20.4	295	3	Q94083	HYPOTHETICAL 33.2 KD P	3.62e-01
17	79	19.8	212	2	O34644	TRIACYLGLYCEROL LIPASE	7.80e-01
18	79	19.8	311	1	Q9YFL3	311AA LONG HYPOTHETICA	7.80e-01
19	79	19.8	471	2	Q9XD57	ACYL-COA LIGASE (FRAGM	7.80e-01
20	79	19.8	706	2	Q47462	PLASMID DNA FOR ENEC-H	7.80e-01
21	79	19.8	7[illegible]	2	[illegible]	HEMOLYSIN [illegible]	[illegible]
22	79	19.8	8[illegible]	14	[illegible]	ENVELOPE GL[illegible]	[illegible]
23	78	19.6	434	10	Q38B91	PREDICTED [illegible]	[illegible]
24	78	19.6	582	2	[illegible]	OXALYL-COA [illegible]	[illegible]
25	77	19.3	398	8	[illegible]	NADH DEHYDR[illegible]	[illegible]
26	77	19.3	5[illegible]	2	[illegible]	NUOM, NADH [illegible]	[illegible]
27	77	19.3	17[illegible]	10	[illegible]	ENHPI [illegible]	[illegible]
28	75	18.8	[illegible]	10	[illegible]	I2P11.12 PR[illegible]	[illegible]
29	75	18.8	425	2	O07714	RIBA.	[illegible]
30	75	18.8	535	14	Q66185	(STRAIN VI[illegible]	[illegible]
31	75	18.8	564	2	Q46523	FIMN.	[illegible]
32	75	18.8	1256	5	Q24022	FLIGHTLESS	[illegible]
33	75	18.8	1[illegible]	4	[illegible]	KIAA0[illegible]	[illegible]
34	74	18.6	232	5	[illegible]	SLA-DP1 BE[illegible]	[illegible]
35	74	18.6	266	7	Q31071	SLA-DR BETA	[illegible]
36	74	18.6	266	7	Q31072	SLA-DR BETA	[illegible]
37	74	18.6	359	2	[illegible]	RSTA2.	[illegible]
38	74	18.6	365	2	[illegible]	RSTA.	[illegible]
39	74	18.6	442	4	[illegible]	RETINAL PA[illegible]	[illegible]
40	73	18.3	222	5	[illegible]	MEMBRANE P[illegible]	[illegible]
41	73	18.3	342	10	Q22713	F8A5.24 PR[illegible]	[illegible]
42	73	18.3	638	10	Q80004	T19L18.6 PR[illegible]	[illegible]
43	73	18.3	1133	15	[illegible]	F2P9.11 PR[illegible]	[illegible]
44	72	18.1	1[illegible]	14	[illegible]	BETA 2 MICR[illegible]	[illegible]
45	72	18.1	774	4	[illegible]	ALPHA [illegible]	[illegible]

ALIGNMENTS

RESULT 1
ID O74544 PRELIMINARY; PRT; 314 AA.
AC O74544;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE YEAST GTR2 HOMOLOG, NOVEL SMALL GTPASE SUBFAMILY
GN SPCC777.05.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Archiascomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae
OC Schizosaccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA SEEGER K., HARRIS D., M[illegible] V., RAJANDREAM M.A., [illegible]
RL Submitted (AUG 1998) to the EMBL/GenBank/DDBJ data
DR EMBL; AL031542; CAA20709.1; -.
SQ SEQUENCE 314 AA; 35691 MW; 34BE04BC CRC32;

Query Match 22.6%; Score 90; DB 3; Length
Best Local Similarity 45.7%; Pred. No. 9.93e-03;
Matches 17; Conservative 6; Mismatches 11;

Db 269 YVAIDSSPVDVQSYEQSDFIDVIIIIGSIYGRSSQL 245
[illegible]
Qy 6 YIGRNSSRV[illegible]

RESULT 2
ID Q66188 PRELIMINARY; PRT; 535 AA.
AC Q66188;
DT 01-NOV-1996 (TrEMBLrel. 02, Created)
DT 01-NOV-1996 (TrEMBLrel. 02, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE (STRAIN D1466) PEPLOMER PROTEIN GENE ENCODING H
DE SUBUNITS, COMPLETE CDS PRECURSOR (FRAGMENT).
OS avian infectious bronchitis virus.
OC Viruses; ssRNA positive-strand viruses, no DNA st
OC Coronaviridae; Coronavirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89163257.
RA KUSTERS J.G., NIESTERS H.G.M., LENSTRA J.A., HORZ

```
RA   DER ZEIJST B.A.M.;
RT   "Phylogeny of antigenic variants of avian coronavirus IBV.",
RL   Virology 169:217-221(1989).
DR   EMBL; M21971; AAA46229.1; -.
DR   PFAM; PF01600; Corona_S1; 1.
KW   Signal.
FT   SIGNAL        1      4       POTENTIAL.
FT   CHAIN         5    >176      POTENTIAL.
FT   CHAIN       179    >535      POTENTIAL.
FT   NON_TER     535    535
SQ   SEQUENCE   535 AA;  59092 MW;  76290200 CRC32;

  Query Match           21.6%;  Score 86;  DB 14;  Length 535;
  Best Local Similarity  26.7%;  Pred. No. 5.05e-02;
  Matches   12;  Conservative   15;  Mismatches  16;  Indels   2;  Gaps   2;

Db      45 AYLVTNETDI-SYNGVSCTVGTIKCGIVINESA-ISFVTKTPIAW 87
           :||  | : : :|:|  | ::| : :| :  :|  :|:   :|: |
Qy       5 GYLSRNSSRVCNYHGCSSAIGILEAGQIYGNSNQLSLSMDSPVEW 49


RESULT    3
ID   Q87628      PRELIMINARY;      PRT;   758 AA.
AC   Q87628;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   GAG (FRAGMENT).
GN   ENV.
OS   Chimpanzee immunodeficiency virus (SIV(cpz)) (CIV).
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96295434.
RA   VANDEN HAESEVELDE M.M., PEETERS M., JANNES G., JANSSENS W., GROEN G.,
RA   SHARP P.M., SAMAN E.;
RT   "Sequence analysis of a highly divergent HIV 1 related lentivirus
RT   isolated from a wild captured chimpanzee.";
RL   Virology 221:346-350(1996).
DR   EMBL; U42720; AAB47730.1; -.
DR   PFAM; PF00516; GP120; 1.
DR   PFAM; PF00517; GP41; 1.
FT   NON_TER     758    758
SQ   SEQUENCE   758 AA;  86475 MW;  D379CEE6 CRC32;

  Query Match           21.6%;  Score 86;  DB 14;  Length 758;
  Best Local Similarity  36.4%;  Pred. No. 5.05e-02;
  Matches   12;  Conservative   10;  Mismatches  10;  Indels   1;  Gaps   1;

Db     438 SCTSSITGIMLEGQIYNETVKVSPAARVADQWR 470
           :|:|:| ||:  |||| ::  :| :   : :||
Qy      19 GCSSAI-GILEAGQIYGNSNQLSLSMDSPVEWR 50


RESULT    4
ID   O12002      PRELIMINARY;      PRT;   858 AA.
AC   O12002;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   ENVELOPE GLYCOPROTEIN GP160.
GN   ENV.
OS   Chimpanzee immunodeficiency virus (SIV(cpz)) (CIV).
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ANT1;
RA   NYAMBI P., LEWI P., PEETERS M., JANSSENS W., HEYNDRICKX L.,
RA   FRANSEN K., ANDRIES K., DEN HAESEVELDE M., HEENEY J., PIOT P.,
RA   DER GROEN G.;
RL   J. Virol. 0:0-0(0).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ANT1;
RA   BLOUIN J.C.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U84094; AAB58529.1; -.
DR   PFAM; PF00516; GP120; 1.
DR   PFAM; PF00517; GP41; 1.
KW   Envelope protein.
SQ   SEQUENCE   858 AA;  98222 MW;  126F6E3C CRC32;

  Query Match           21.6%;  Score 86;  DB 14;  Length 858;
  Best Local Similarity  36.4%;  Pred. No. 5.05e-02;
  Matches   12;  Conservative   10;  Mismatches  10;  Indels   1;  Gaps   1;

Db     438 SCTSSITGIMLEGQIYNETVKVSPAARVADQWR 470
           :|:|:| ||:  |||| ::  :| :   : :||
Qy      19 GCSSAI-GILEAGQIYGNSNQLSLSMDSPVEWR 50


RESULT    5
ID   O12003      PRELIMINARY;      PRT;   874 AA.
AC   O12003;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   ENVELOPE GLYCOPROTEIN GP160.
GN   ENV.
OS   Chimpanzee immunodeficiency virus (SIV(cpz)) (CIV).
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ANT2;
RA   NYAMBI P., LEWI P., PEETERS M., JANSSENS W., HEYNDRICKX L.,
RA   FRANSEN K., ANDRIES K., DEN HAESEVELDE M., HEENEY J., PIOT P.,
RA   DER GROEN G.;
RL   J. Virol. 0:0-0(0).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ANT2;
RA   BLOUIN J.C.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U84095; AAB58530.1; -.
DR   PFAM; PF00516; GP120; 1.
DR   PFAM; PF00517; GP41; 1.
KW   Envelope protein.
SQ   SEQUENCE   874 AA;  100143 MW;  24719EE9 CRC32;

  Query Match           21.6%;  Score 86;  DB 14;  Length 874;
  Best Local Similarity  36.4%;  Pred. No. 5.05e-02;
  Matches   12;  Conservative   10;  Mismatches  10;  Indels   1;  Gaps   1;

Db     453 SCTSSITGIMLEGQIYNETVKVSPAARVADQWR 485
           :|:|:| ||:  |||| ::  :| :   : :||
Qy      19 GCSSAI-GILEAGQIYGNSNQLSLSMDSPVEWR 50


RESULT    6
ID   Q40996      PRELIMINARY;      PRT;   171 AA.
AC   Q40996;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   ALBUMIN 2 (FRAGMENT).
OS   Pinus strobus (Eastern white pine).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Coniferopsida; Coniferales; Pinaceae;
OC   Pinus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PROTHALLIUM;
RA   RICE J., KAMALAY J.C.;
RL   Submitted (OCT-1991) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
```

```
RC   TISSUE=PROTHALLIUM;
RA   RICE J., KAMALAY J.C.;
RL   Submitted (OCT-1991) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X62434; CAA44299.1; -.
DR   MENDEL; 12151; Pinst;1310;12151.
DR   PFAM; PF00234; tryp_alpha_amyl; 1.
FT   NON_TER       1       1
SQ   SEQUENCE   171 AA;  20070 MW;  FBCDCA8F CRC32;

  Query Match           21.4%;  Score 85;  DB 10;  Length 171;
  Best Local Similarity  27.8%;  Pred. No. 7.54e-02;
  Matches   10;  Conservative   14;  Mismatches  10;  Indels   2;  Gaps   2;

Db      91 CRAIQQVLDQSHSYDSITEDLSMDSDAALNQRRRRR 126
           |:  :|: :: | :: : ||:  |::: ::|||
Qy      20 CSSAIGILEAGQIYGNSNQ-LSLSMDSPV-EWRRRR 53


RESULT    7
ID   O12004      PRELIMINARY;      PRT;   862 AA.
AC   O12004;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   ENVELOPE GLYCOPROTEIN GP160.
GN   ENV.
OS   Chimpanzee immunodeficiency virus (SIV(cpz)) (CIV).
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ANT3;
RA   NYAMBI P., LEWI P., PEETERS M., JANSSENS W., HEYNDRICKX L.,
RA   FRANSEN K., ANDRIES K., DEN HAESEVELDE M., HEENEY J., PLOT P.,
RA   DER GROEN G.;
RL   J. Virol. 0:0-0(0).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ANT3;
RA   BLOUIN J.C.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U84096; AAB58531.1; -.
DR   PFAM; PF00516; GP120; 1.
DR   PFAM; PF00517; GP41; 1.
KW   Envelope protein.
SQ   SEQUENCE   862 AA;  99009 MW;  D5BAC67F CRC32;

  Query Match           21.4%;  Score 85;  DB 14;  Length 862;
  Best Local Similarity  36.4%;  Pred. No. 7.54e-02;
  Matches   12;  Conservative   10;  Mismatches  10;  Indels   1;  Gaps   1;

Db     442 SCTSSITGIMLEGQIYNENVNVSPAARVADQWR 474
           :|:|:| ||:  ||| ::  :| :   : :||
Qy      19 GCSSAI GILEAGQIYGNSNQLSLSMDSPVEWR 50


RESULT    8
ID   O12011      PRELIMINARY;      PRT;   870 AA.
AC   O12011;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   ENVELOPE GLYCOPROTEIN GP160.
GN   ENV.
OS   Chimpanzee immunodeficiency virus (SIV(cpz)) (CIV).
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ANT10;
RA   NYAMBI P., LEWI P., PEETERS M., JANSSENS W., HEYNDRICKX L.,
RA   FRANSEN K., ANDRIES K., DEN HAESEVELDE M., HEENEY J., PLOT P.,
RA   DER GROEN G.;
RL   J. Virol. 0:0-0(0).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ANT10;
RA   BLOUIN J.C.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ dat
DR   EMBL; U84103; AAB58508.1;
DR   PFAM; PF00516; GP120; 1.
DR   PFAM; PF00517; GP41; 1.
KW   Envelope protein.
SQ   SEQUENCE   870 AA;  99737 MW;  B5999B3B CRC32;

  Query Match           21.4%;  Score 85;  DB 14;  Le
  Best Local Similarity  36.4%;  Pred. No. 7.54e-02;
  Matches   12;  Conservative   10;  Mismatches  10;

Db     451 SCTSSITGIMLEGQIYINVNVSPAARVADQWR 483
           :|:|:| ||:  || | ::  :: :   : :|
Qy      19 GCSSAI-GILEAGQIYGNSNQLSLSMDSPVEWR 50


RESULT    9
ID   O12010      PRELIMINARY;      PRT;   862 AA.
AC   O12010;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   ENVELOPE GLYCOPROTEIN GP160.
GN   ENV.
OS   Chimpanzee immunodeficiency virus (SIV(cpz)) (CIV).
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ANT9;
RA   NYAMBI P., LEWI P., PEETERS M., JANSSENS W., HEYNDRICKX L.,
RA   FRANSEN K., ANDRIES K., DEN HAESEVELDE M., HEENEY J., PLOT P.,
RA   DER GROEN G.;
RL   J. Virol. 0:0-0(0).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ANT9;
RA   BLOUIN J.C.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ dat
DR   EMBL; U84102; AAB58507.1;
DR   PFAM; PF00516; GP120; 1.
DR   PFAM; PF00517; GP41; 1.
KW   Envelope protein.
SQ   SEQUENCE   862 AA;  98799 MW;  E727ED41 CRC32;

  Query Match           21.4%;  Score 85;  DB 14;  Le
  Best Local Similarity  36.4%;  Pred. No. 7.54e-02;
  Matches   12;  Conservative   10;  Mismatches  10;

Db     443 SCTSSITGIMLEGQIYINVNVSPAARVADQWR 475
           :|:|:| ||:  ||| ::  :: :   : :|
Qy      19 GCSSAI-GILEAGQIYGNSNQLSLSMDSPVEWR 50


RESULT   10
ID   O12008      PRELIMINARY;      PRT;   854 AA.
AC   O12008;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   ENVELOPE GLYCOPROTEIN GP160.
GN   ENV.
OS   Chimpanzee immunodeficiency virus (SIV(cpz)) (CIV).
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ANT7;
RA   NYAMBI P., LEWI P., PEETERS M., JANSSENS W., HEYN
RA   FRANSEN K., ANDRIES K., DEN HAESEVELDE M., HEENEY
RA   DER GROEN G.;
RL   J. Virol. 0:0-0(0).
```

```
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ANT7;
RA   BLOUIN J.C.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U84100; AAB58535.1; -.
DR   PFAM; PF00516; GP120; 1.
DR   PFAM; PF00517; GP41; 1.
KW   Envelope protein.
SQ   SEQUENCE   854 AA;  97975 MW;  70577AF3 CRC32;

  Query Match           20.9%;  Score 83;  DB 14;  Length 854;
  Best Local Similarity  36.4%;  Pred. No. 1.66e-01;
  Matches   12;  Conservative   10;  Mismatches  10;  Indels   1;  Gaps   1;

Db    435 SCTSSITGIMLEGQIYNENVSVSPAARVADQWR 467
          :|:|:| ||:  |||| ::  :| :   : :||
Qy     19 GCSSAI-GILEAGQIYGNSNQLSLSMDSPVEWR 50


RESULT   11
ID   O12006      PRELIMINARY;      PRT;   866 AA.
AC   O12006;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   ENVELOPE GLYCOPROTEIN GP160.
GN   ENV.
OS   Chimpanzee immunodeficiency virus (SIV(cpz)) (CIV).
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ANT5;
RA   NYAMBI P., LEWI P., PEETERS M., JANSSENS W., HEYNDRICKX L.,
RA   FRANSEN K., ANDRIES K., DEN HAESEVELDE M., HEENEY J., PIOT P.,
RA   DER GROEN G.;
RL   J. Virol. 0:0-0(0).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ANT5;
RA   BLOUIN J.C.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U84098; AAB58533.1; -.
DR   PFAM; PF00516; GP120; 1.
DR   PFAM; PF00517; GP41; 1.
KW   Envelope protein.
SQ   SEQUENCE   866 AA;  99464 MW;  10CF2951 CRC32;

  Query Match           20.9%;  Score 83;  DB 14;  Length 866;
  Best Local Similarity  36.4%;  Pred. No. 1.66e-01;
  Matches   12;  Conservative   10;  Mismatches  10;  Indels   1;  Gaps   1;

Db    446 SCTSSITGIMLEGQIYNENVSVSPAARVADQWR 478
          :|:|:| ||:  |||| ::  :| :   : :||
Qy     19 GCSSAI-GILEAGQIYGNSNQLSLSMDSPVEWR 50


RESULT   12
ID   O12009      PRELIMINARY;      PRT;   868 AA.
AC   O12009;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   ENVELOPE GLYCOPROTEIN GP160
GN   ENV.
OS   Chimpanzee immunodeficiency virus (SIV(cpz)) (CIV)
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ANT8;
RA   NYAMBI P., LEWI P., PEETERS M., JANSSENS W., HEYNDRICKX L.,
RA   FRANSEN K., ANDRIES K., DEN HAESEVELDE M., HEENEY J., PIOT P.,
RA   DER GROEN G.;
RL   J. Virol. 0:0-0(0).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ANT8;
RA   BLOUIN J.C.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U84101; AAB58536.1; -.
DR   PFAM; PF00516; GP120; 1.
DR   PFAM; PF00517; GP41; 1.
KW   Envelope protein.
SQ   SEQUENCE   868 AA;  99508 MW;  DA13BEDE CRC32;

  Query Match           20.9%;  Score 83;  DB 14;  Length 868;
  Best Local Similarity  36.4%;  Pred. No. 1.66e-01;
  Matches   12;  Conservative   10;  Mismatches  10;  Indels   1;  Gaps   1;

Db    449 SCTSSITGIMLEGQIYNENVSVSPAARVADQWR 481
          :|:|:| ||:  |||| ::  :| :   : :||
Qy     19 GCSSAI-GILEAGQIYGNSNQLSLSMDSPVEWR 50


RESULT   13
ID   Q9XZY8      PRELIMINARY;      PRT;  2081 AA.
AC   Q9XZY8;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   L302.11 PROTEIN.
GN   L302.11.
OS   Leishmania major.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   MURRAY B., QUAIL M., LAWSON D., HARRIS D., RAJANDREAM M., IVENS A.,
RA   BARRELL B.;
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL031909; CAB42353.1; -.
SQ   SEQUENCE   2081 AA;  224389 MW;  F3037DDE CRC32;

  Query Match           20.9%;  Score 83;  DB 5;  Length 2081;
  Best Local Similarity  37.1%;  Pred. No. 1.66e-01;
  Matches   13;  Conservative   9;  Mismatches  11;  Indels   2;  Gaps   2;

Db    778 HGCASG-SRADGGAAAGNAAALHVDAMSS-DWRRR 810
          |||:|: :  ::|   ||:  | :   |: :||||
Qy     18 HGCSSAICILEAGQIYGNSNQLSLSMDSPVEWRRR 52


RESULT   14
ID   Q9YBJ9      PRELIMINARY;      PRT;   683 AA.
AC   Q9YBJ9;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   683AA LONG HYPOTHETICAL PROTEIN.
GN   APE1599.
OS   Aeropyrum pernix.
OC   Archaea; Crenarchaeota; Aeropyrum.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K1;
RX   MEDLINE; 99310339.
RA   KAWARABAYASI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y.,
RA   JIN-NO K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,
RA   HOSOYAMA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,
RA   TAKAMIYA M., MASUDA S., FUNAHASHI T., TANAKA T., KUDOH Y.,
RA   YAMAZAKI J., KUSHIDA N., OGURA A., ARAI K., KUBOTA K., NAKAMURA Y.,
RA   NOMURA N., SAKO Y., KIKUCHI H.;
RT   "Complete genome sequence of an aerobic hyper-thermophilic
RT   crenarchaeon, Aeropyrum pernix K1.";
RL   DNA Res. 6:83-101(1999).
DR   EMBL; AP000062; BAA80599.1; -.
SQ   SEQUENCE   683 AA;  75229 MW;  4447EC52 CRC32;
```

```
  Query Match           20.6%;  Score 82;  DB 1;  Length 683;
  Best Local Similarity  36.4%;  Pred. No. 2.46e-01;
  Matches   12;  Conservative    8;  Mismatches  12;  Indels   1;  Gaps   1;

Db    631 SEVQVLRLGGYVLVSSEVLRLAEEMRAEWRRRR 663
          | : :|  :| :  :|: | |: :   |||||| 
Qy     22 SAIGILE-AGQIYGNSNQLSLSMDSPVEWRRRR 53


RESULT   15
ID   O12005       PRELIMINARY;      PRT;   866 AA.
AC   O12005;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   ENVELOPE GLYCOPROTEIN GP160.
GN   ENV.
OS   Chimpanzee immunodeficiency virus (SIV(cpz)) (CIV).
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ANT4;
RA   NYAMBI P., LEWI P., PEETERS M., JANSSENS W., HEYNDRICKX L.,
RA   FRANSEN K., ANDRIES K., DEN HAESEVELDE M., HEENEY J., PIOT P.,
RA   DER GROEN G.;
RL   J. Virol. 0:0-0(0).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ANT4;
RA   BIOUIN J.C.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U84097; AAB58542.1; -.
DR   PFAM; PF00516; GP120; 1.
DR   PFAM; PF00517; GP41; 1.
KW   Envelope protein.
SQ   SEQUENCE   866 AA;  99373 MW;  ACE2DB2D CRC32;

  Query Match           20.6%;  Score 82;  DB 14;  Length 866;
  Best Local Similarity  36.4%;  Pred. No. 2.46e-01;
  Matches   12;  Conservative    9;  Mismatches  11;  Indels   1;  Gaps   1;

Db    445 SCISSIIGIMLEGQIYNQNVSVSPAARVADQWR 477
          :|:|:| ||:  |:|| :  :| :   : :||
Qy     19 GCSSAI-GILEAGQIYGNSNQLSLSMDSPVEWR 50


Search completed: Fri Aug 25 18:09:11 2000
Job time : 40 secs.
```

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 25 18:03:56 2000; MasPar time 4.74 Seconds
338.281 Million cell updates/sec

Tabular output not generated.

Title: >US-09-257-585-7
Description: (1-34) from US09257585.pep
Perfect Score: 250
Sequence: 1 MRPAIPTQLYFPTDRSFLAILHQLTLTTMSLVDQ 34

Scoring table: PAM 150
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir64
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 34.439; Variance 63.944; scale 0.539

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	94	37.6	645	2	T11127	NADH dehydrogenase su	1.04e-02
2	92	36.8	399	2	D42586	glucuronosyltransfera	6.53e-01
3	82	32.8	431	1	JX0364	antithrombin III - pi	6.53e-01
4	81	32.4	352	1	A45920	urate oxidase (EC 1.7	9.08e-01
5	80	32.0	328	1	DEHUE7	estradiol 17beta-dehy	1.26e+00
6	80	32.0	465	2	I59611	antithrombin III - mo	1.26e+00
7	79	31.6	422	2	G72643	probable transketolas	1.75e+00
8	77	30.8	144	2	G71205	hypothetical protein	3.32e+00
9	77	30.8	318	2	T11441	NADH dehydrogenase (u	3.32e+00
10	77	30.8	466	2	A25330	bacteriochlorophyll a	3.32e+00
11	77	30.8	464	1	XHHU3	antithrombin III prec	3.32e+00
12	76	30.4	342	1	S29133	urate oxidase (EC 1.7	4.57e+00
13	75	30.0	479	2	A54077	ubiquinol--cytochrome	6.26e+00
14	74	29.6	242	2	D71542	probable ribose 5 pho	8.55e+00
15	74	29.6	522	2	S68362	glucose transport pro	8.55e+00
16	74	29.6	522	2	A31556	glucose transport pro	8.55e+00
17	74	29.6	567	1	ACFFA1	nicotinic acetylcholi	8.55e+00
18	74	29.6	691	2	A34883	band 4.2 protein - hu	8.55e+00
19	74	29.6	721	2	A34865	protein 4.2 - human	8.55e+00
20	73	29.2	282	2	E71543	hypothetical protein	1.16e+01
21	73	29.2	479	1	S17408	ubiquinol--cytochrome	1.16e+01
22	73	29.2	465	1	S28219	antithrombin III prec	1.16e+01
23	73	29.2	523	2	S06920	glucose transport pro	1.16e+01
24	72	28.8	248	2	S76774	thiolase-like	[illegible]
25	72	28.8	344	2	S42652	estradiol	[illegible]
26	72	28.8	344	2	S47901	estradiol	[illegible]
27	72	28.8	[illegible]	1	A61455	antithrombin	[illegible]
28	72	28.8	610	2	H71612	asparagine	[illegible]
29	72	28.8	670	2	B43655	glycine-rich	[illegible]
30	71	28.4	108	2	S25808	hypothetic	[illegible]
31	71	28.4	279	2	A71499	hypothetic	[illegible]
32	71	28.4	333	1	JC5166	ketol-acid	[illegible]
33	71	28.4	354	2	S51143	PMO-protei	[illegible]
34	71	28.4	379	2	S48156	ubiquinol	[illegible]
35	71	28.4	496	2	E70622	probable	[illegible]
36	71	28.4	601	2	T11453	NADH5 prot	[illegible]
37	71	28.4	749	2	B49604	two-compo	[illegible]
38	71	28.4	759	2	A45848	competence	[illegible]
39	71	28.4	901	2	I11156	hypothetic	[illegible]
40	71	28.4	3085	2	I70327	polyprotei	[illegible]
41	70	28.0	223	2	I75815	sigma fact	[illegible]
42	70	28.0	262	2	I-5825	pectin ac	[illegible]
43	70	28.0	308	2	S22919	ubiquinol	[illegible]
44	70	28.0	346	1	S29763	urate oxi	[illegible]
45	70	28.0	1322	2	B71443	hypotheti	[illegible]

ALIGNMENTS

RESULT 1
ENTRY T11137 #type complete
TITLE NADH dehydrogenase subunit 5 - acorn [illegible]
ORGANISM #formal_name Schizochytrium balanoi [illegible]
DATE 16-Jul-1999 #sequence_revision 16-Jul-1999 [illegible]
20-Sep-1999
ACCESSIONS T11137
REFERENCE Z17250
#authors Castresana, J.; Feldmaier-Fuchs, G.; S [illegible]
Paabo, S.
#journal Genetics (1998) 150:1115-1123
#title The mitochondrial genome of the hemich [illegible]
carnosus and the evolution of deuter [illegible]
#cross-references MUID:98461665
#accession T11137
##status preliminary; translated from GP [illegible]
##molecule_type DNA
##residues 1-645 ##label 'AS
##cross-references EMBL:AF [illegible]; NID:g [illegible];
PIDN:AAC11645.1
GENETICS
#genome mitochondrion
CLASSIFICATION #superfamily NADH dehydrogenase (ubiq [illegible]
KEYWORDS mitochondrion
SUMMARY #length 645 #molecular-weight 69454 [illegible]

Query Match 37.6%; Score 94; DB 2; Leng [illegible]
Best Local Similarity 35.7%; Pred. No. 1.04e-02;
Matches 17; Conservative 5; Mismatches 4; [illegible]

Db 45 PLICAQLKF-TNR-PLSIIH_ISPTIISLV 72
[illegible]
Qy 3 PAIPTQLYFPTDRSFLAILH_LTLTTMSLV 32

RESULT 2
ENTRY D42586 #type fragment
TITLE glucuronosyltransferase (EC 2.4.1.17, [illegible]
(fragment)
ORGANISM #formal_name Homo sapiens #common_nam [illegible]
DATE 04-Mar-1994 #sequence_revision 04-Mar [illegible]
29-Sep-1994
ACCESSIONS D42586
REFERENCE A42586
#authors Ritter, J.K.; Chen, F.; Sheen, Y.Y.; [illegible]
Yeatman, M.L.; Owens, I.S.
#journal J. Biol. Chem. (1992) 267:3257-3261

```
   #title          A novel complex locus UGT1 encodes human bilirubin, phenol,
                     and other UDP-glucuronosyltransferase isozymes with
                     identical carboxyl termini.
   #cross-references MUID:92147680
   #accession      D42586
      ##status          preliminary; not compared with conceptual translation
      ##molecule_type nucleic acid
      ##residues       1-289 ##label RTF
      ##cross-references GB:M84127; NID:g340134; PIDN:AAA02020.1; PID:g340135
      ##note           sequence extracted from NCBI backbone (NCBIP:81436)
CLASSIFICATION    #superfamily glucuronosyltransferase
KEYWORDS          glycosyltransferase; hexosyltransferase
SUMMARY           #length 289  #checksum 4342

  Query Match             42.8%;  Score 82;  DB 2;  Length 289;
  Best Local Similarity   46.4%;  Pred. No. 6.53e-01;
  Matches    13;  Conservative    7;  Mismatches    5;  Indels    3;  Gaps    3;

Db      97 TQLYFPTDRSFLA-ILHQLT-LTTMSLV 127
           ||||| ||: || ::: :: | ||||
Qy       7 TQLYFPTDRSFLA-ILHQLT-LTTMSLV 32


RESULT    3
ENTRY             JX0364      #type complete
TITLE             antithrombin III - pig
ORGANISM          #formal_name Sus scrofa domestica #common_name domestic pig
DATE              10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change
                    10-Sep-1999
ACCESSIONS        JX0364
REFERENCE         JX0364
   #authors       Tokunaga, F.; Goto, T.; Wakabayashi, S.; Koide, T.
   #journal       J. Biochem. (1994) 116:1164-1170
   #title         Amino acid sequence of porcine antithrombin III.
   #cross-references MUID:95204393
   #accession     JX0364
      ##molecule_type protein
      ##residues       1-431 ##label IOK
FUNCTION
   #description   inhibits in blood plasma thrombin and activated coagulation
                    factor X, either weakly alone or strongly in the presence
                    of heparin
CLASSIFICATION    #superfamily antithrombin III
KEYWORDS          glycoprotein; heparin binding
FEATURE
   378-382             #region hinge #status predicted\
   9-129,22-96,
   218-431             #disulfide_bonds #status predicted\
   136,156,193         #binding_site carbohydrate (Asn) (covalent) #status
                         experimental\
   394                 #inhibitory_site Arg (thrombin) #status predicted
SUMMARY           #length 431  #molecular-weight 48936  #checksum 362

  Query Match             42.8%;  Score 82;  DB 1;  Length 431;
  Best Local Similarity   24.2%;  Pred. No. 6.53e-01;
  Matches    8;  Conservative   17;  Mismatches    8;  Indels    0;  Gaps    0;

Db     394 RSLNPARVTPKANRPFLVLIREVALNTIIFMGR 426
           |:  |::: | ::|:|| :::::: |: :: :
Qy       2 RPAIPTQLYFPTDRSFLAILHQLTLTTMSLVDQ 34


RESULT    4
ENTRY             A35920      #type complete
TITLE             urate oxidase (EC 1.7.3.3) - fruit fly (Drosophila
                    melanogaster)
ALTERNATE_NAMES   uricase
ORGANISM          #formal_name Drosophila melanogaster
DATE              09-Nov-1990 #sequence_revision 27-Jun-1994 #text_change
                    14-Aug-1999
ACCESSIONS        A35920; S08676
REFERENCE         A35920
   #authors       Wallrath, L.L.; Burnett, J.B.; Friedman, T.B.
   #journal       Mol. Cell. Biol. (1990) 10:5114-5127
   #title         Molecular characterization of the Drosophila melanogaster
                    urate oxidase gene, an ecdysone repressible gene expressed
                    only in the malpighian tubules.
   #cross-references MUID:90377201
   #accession     A35920
      ##molecule_type DNA
      ##residues       1-352 ##label WAL
      ##cross-references EMBL:X51940; NID:g8798; PIDN:CAA36203.1; PID:g6799
GENETICS
   #gene          UO
      ##cross-references FlyBase:FBgn0003961
   #introns        192/1
CLASSIFICATION    #superfamily urate oxidase
KEYWORDS          copper binding, oxidoreductase, peroxisome
FEATURE
   350-352             #region peroxisome/SKL, C-terminal targeting signal
                         (S-[RKH]-L) motif\
   170,172             #binding_site copper (his) (type 2) #status predicted
SUMMARY           #length 352  #molecular-weight 40034  #checksum 5711

  Query Match             32.4%;  Score 81;  DB 1;  Length 352;
  Best Local Similarity   27.6%;  Pred. No. 9.08e-01;
  Matches     8;  Conservative   12;  Mismatches    8;  Indels    1;  Gaps    1;

Db     277 PSVQHTLYL-SERQVLDVLPQVSVISMIM 304
           |::   ||: ::|  | :| |::: :|::
Qy       3 PAIPTQLYFPTDRSFLAILHQLTLTTMSL 31


RESULT    5
ENTRY             DEHUE7      #type complete
TITLE             estradiol 17beta-dehydrogenase (EC 1.1.1.62) type 1 - human
ALTERNATE_NAMES   17beta-estradiol dehydrogenase; 17beta hydroxysteroid
                    dehydrogenase
ORGANISM          #formal_name Homo sapiens #common_name man
DATE              31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change
                    24-Sep-1999
ACCESSIONS        A36081; S29288; S01654; S43212; A40146, A60440, A21031;
                    A60628; G02067, A37190, A37239
REFERENCE         A36081
   #authors       Luu-The, V.; Labrie, C.; Simard, J.; Lachance, Y.; Zhao,
                    H.F.; Couet, J.; Leblanc, G.; Labrie, F.
   #journal       Mol. Endocrinol. (1990) 4:268-275
   #title         Structure of two in tandem human 17beta-hydroxysteroid
                    dehydrogenase genes.
   #cross-references MUID:90231340
   #accession     A36081
      ##molecule_type DNA
      ##residues       1-328 ##label LUU
      ##cross-references GB:M27138; NID:g181950; PIDN:AAB16941.1; PID:g181951
REFERENCE         S29288
   #authors       Peltoketo, H.; Isomaa, V.; Vihko, R.
   #journal       Eur. J. Biochem. (1992) 209:459-466
   #title         Genomic organization and DNA sequences of human
                    17beta-hydroxysteroid dehydrogenase genes and flanking
                    regions. Localization of multiple Alu sequences and
                    putative cis-acting elements.
   #cross-references MUID:93011163
   #accession     S29288
      ##molecule_type DNA
      ##residues       1-328 ##label PEL
      ##cross-references EMBL:M84472; NID:g806392; PIDN:AAB16942.1;
                          PID:g177127
      ##note          the authors did not translate the codon for residue 1
      ##note          the authors translated the codon AAC for residue 243 as
                       Asp
REFERENCE         S01654
   #authors       Peltoketo, H., Isomaa, V., Maentausta, O.; Vihko, R.
   #journal       FEBS Lett. (1988) 239:73-77
   #title         Complete amino acid sequence of human placental
                    17 beta-hydroxysteroid dehydrogenase deduced from cDNA.
   #cross-references MUID:89031223
```

```
    #accession     S01654
       ##molecule_type mRNA
       ##residues       2-328 ##label PEL2
       ##cross-references GB:X13440; NID:g23364; PIDN:CAA31792.1; PID:g23365
    #accession     S43212
       ##molecule_type protein
       ##residues       2-24;53-68;205 223 ##label PE2
REFERENCE          A40146
    #authors       Luu-The, V.; Labrie, C.; Zhao, H.F.; Couet, J.; Lachance, Y.;
                     Simard, J.; Leblanc, G.; Cote, J.; Berube, D.; Gagne, R.;
                     Labrie, F.
    #journal       Mol. Endocrinol. (1989) 3:1301-1309
    #title         Characterization of cDNAs for human estradiol
                     17beta-dehydrogenase and assignment of the gene to
                     chromosome 17: evidence of two mRNA species with distinct
                     5'-termini in human placenta.
    #cross-references MUID:89384667
    #accession     A40146
       ##molecule_type mRNA
       ##residues       1-328 ##label THE
       ##cross-references GB:M27138; NID:g181950; PIDN:AAA16941.1; PID:g181951
REFERENCE          A60440
    #authors       Gast, M.J.; Sims, H.F.; Murdock, G.L.; Gast, P.M.; Strauss,
                     A.W.
    #journal       Am. J. Obstet. Gynecol. (1989) 161:1726-1731
    #title         Isolation and sequencing of a complementary deoxyribonucleic
                     acid clone encoding human placental 17beta-estradiol
                     dehydrogenase: identification of the putative cofactor
                     binding site.
    #cross-references MUID:90102621
    #accession     A60440
       ##status        not compared with conceptual translation
       ##molecule_type mRNA
       ##residues       5-68,'G',70-91,'EAWACAAG',102-158,160-178,'F',181-311,
                        'A',313-328 ##label GAS
REFERENCE          A21031
    #authors       Murdock, G.L.; Chin, C.C.; Offord, R.E.; Bradshaw, R.A.;
                     Warren, J.C.
    #journal       J. Biol. Chem. (1983) 258:11460-11464
    #title         Human placental estradiol 17-beta-dehydrogenase.
                     Identification of a single histidine residue
                     affinity-labeled by both 3-bromoacetoxyestrone and
                     12-beta-bromoacetoxy-4-estrene-3,17-dione.
    #cross-references MUID:84008135
    #accession     A21031
       ##molecule_type protein
       ##residues       220-224 ##label MUR
REFERENCE          A60628
    #authors       Murdock, G.L.; Chin, C.C.; Warren, J.C.
    #journal       Biochemistry (1986) 25:641-646
    #title         Human placental estradiol 17beta-dehydrogenase. sequence of a
                     histidine-bearing peptide in the catalytic region.
    #cross-references MUID:86159758
    #accession     A60628
       ##molecule_type protein
       ##residues       208-215 ##label MU2
REFERENCE          A37190
    #authors       Labrie, F.; Luu-The, V.; Labrie, C.; Berube, D.; Couet, J.;
                     Zhao, H.F.; Gagne, R.; Simard, J.
    #journal       J. Steroid Biochem. (1989) 34:189-197
    #title         Characterization of two mRNA species encoding human estradiol
                     17beta-dehydrogenase and assignment of the gene to
                     chromosome 17.
    #cross-references MUID:90173172
    #contents      annotation
REFERENCE          G09150
    #authors       Shen, Y.
    #submission    submitted to the EMBL Data Library, August 1995
    #accession     G02067
       ##status        translated from GB/EMBL/DDBJ
       ##molecule_type DNA
       ##residues       1-328 ##label SHE
       ##cross-references EMBL:U34879; NID:g975894; PIDN:AAD05019.1;
                         PID:g975895
GENETICS
    #gene          GDB:HSD17B1; Hor:7B.; EDHB
       ##cross-references GenLoc:17q
    #map_position 17q11-q25
    #note          ES1; ovary; placenta; breast cancer
FUNCTION
    #description   catalyzes the reversible oxidation of
                     estrone using NAD(P) as coenzyme
    #note          also catalyzes the reversible oxidatio
                     to dehydroepiandrosterone (DHEA), an
                     4-androstenedione
CLASSIFICATION     #superfamily estradiol 17beta-dehydro
                     alcohol dehydrogenase homology
KEYWORDS           oxidoreductase
FEATURE
    2-328              #product estradiol 17beta-dehydr
                         experimental #label MATX
    4-189              #domain short-chain alcohol dehyd
                         #label SADHX
    211,214,222        #active_site His #status experime
SUMMARY            #length 328 #molecular-weight 34860

  Query Match           42.9%;  Score 80;  DB 1;  Leng
  Best Local Similarity  45.5%;  Pred. No. 1.2e-
  Matches   10;  Conservative    7;  Mismatches   4;

Db     245 LKAPKPTLRYFTIEF-FTPLER 265
           :|:: ||  || |:| ||:::|
Qy       1 MRPAIPTQLYFPTDRSPLAILE 22


RESULT     6
ENTRY              I59611      #type complete
TITLE              antithrombin III - mouse
ORGANISM           #formal_name Mus sp. #common_name mou
DATE               02-Aug-1990 #sequence_revision 02-Aug
                     16-Jul-1999
ACCESSIONS         I59611
REFERENCE          I59611
    #authors       Wu, J.K.; Sheffield, W.P.; Blajchman,
    #journal       Thromb. Haemost. (1992) 68:291-296
    #title         Molecular cloning and cell-free expre
                     antithrombin III.
    #cross-references MUID:93
    #accession     I59611
       ##status        preliminary; translated from GB/
       ##molecule_type mRNA
       ##residues       1-465 ##label KES
       ##cross-references GB:S47225; NID:g258958; PIDN:A
CLASSIFICATION     #superfamily antithrombin III
SUMMARY            #length 465 #molecular-weight 52

  Query Match           42.9%;  Score 80;  DB 2;  Leng
  Best Local Similarity  44.2%;  Pred. No. 1.2e-
  Matches    8;  Conservative    7;  Mismatches   5;

Db     426 KSLNFNRVTTKANRPFIVLIREVALNTIIFMGR 458
           |:  | :: | ::|:|| ::::::| |: :: :
Qy       2 PPAIPTQLYFPTDRSPLAILPLLLLIMSLVIG 34


RESULT     7
ENTRY              G72643      #type complete
TITLE              probable transketolase APE0583 - Aero
ORGANISM           #formal_name Aeropyrum pernix
DATE               20-Aug-1999 #sequence_revision 20-Aug
                     20-Aug-1999
ACCESSIONS         G72643
REFERENCE          A72450
    #authors       Kawarabayasi, Y.; Hino, Y.; Horikawa,
                     Haikawa, Y.; Hirono, K.; Takahashi,
                     S.; Aukai, A.; Kosugi, H.; Hosoyama,
```

```
                    Y.; Nishijima, K.; Nakagawa, H.; Takamiya, M., Masuda, S.,
                    Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.;
                    Kushida, N.; Oguchi, A.; Aoki, K.; Kubota, K.; Nakamura,
                    Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
   #journal         DNA Res. (1999) 6:83-101
   #title           Complete genome sequence of an aerobic hyper-thermophilic
                      Crenarchaeon, Aeropyrum pernix K1.
   #cross-references MUID:99310339
   #accession       G72643
      ##status          preliminary
      ##molecule_type DNA
      ##residues        1-322 ##label KAW
      ##cross-references DDBJ:AP000060; NID:g5104188; PIDN:BAA79551.1;
                        PID:d1043337; PID.g5104236
      ##experimental_source strain K1
GENETICS
   #gene            APE0583
SUMMARY             #length 322  #molecular-weight 34574  #checksum 6060

  Query Match             41.6%;  Score 79;  DB 2;  Length 322;
  Best Local Similarity   24.2%;  Pred. No. 1.75e+00;
  Matches     8;  Conservative    13;  Mismatches  12;  Indels    0;  Gaps    0;

Db     282 RIGVPPGVYGSSSRDYMSLIPHLPLDPESVAAP 314
           | ::|  :| :: | :::|:::| |   |:  :
Qy       2 RPAIPTQLYFPTDRSFLAILHQLTLTTMSLVDQ 34


RESULT    8
ENTRY               G71205      #type complete
TITLE               hypothetical protein PH1913 - Pyrococcus horikoshii
ORGANISM            #formal_name Pyrococcus horikoshii
DATE                14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change
                      14-Aug-1998
ACCESSIONS          G71205
REFERENCE           A71000
   #authors         Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.;
                      Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.;
                      Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.;
                      Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.;
                      Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi,
                      A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.,
                      Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.
   #journal         DNA Res. (1998) 5:55-76
   #title           Complete sequence and gene organization of the genome of a
                      hyper-thermophilic archaebacterium, Pyrococcus horikoshii
                      OT3.
   #cross-references MUID:98344137
   #accession       G71205
      ##status          preliminary; nucleic acid sequence not shown;
                         translation not shown
      ##molecule_type DNA
      ##residues        1-144 ##label KAW
      ##cross-references GB:AP000007; NID:g3236134; PID:d1031981; PID:g3258355
      ##experimental_source strain OT3
      ##note            this accession replaces an interim accession for a
                         sequence replaced by GenBank
GENETICS
   #gene            PH1913
SUMMARY             #length 144  #molecular-weight 15493  #checksum 1694

  Query Match             30.8%;  Score 77;  DB 2;  Length 144;
  Best Local Similarity   50.0%;  Pred. No. 3.32e+00;
  Matches     9;  Conservative     3;  Mismatches   6;  Indels    0;  Gaps    0;

Db     127 LRFTIPSKLYFTLENSFL 144
           :|  ||: |||  : |||
Qy       1 MRPAIPTQLYFPTDRSFL 18


RESULT    9
ENTRY               I11441      #type complete
TITLE               NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 1 -
                      nine banded armadillo mitochondrion (SGC1)
ORGANISM            #formal_name mitochondrion Dasypus novemcinctus #common_name
                      nine-banded armadillo
DATE                16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change
                      13-Aug-1999
ACCESSIONS          I11441
REFERENCE           S17272
   #authors         Arnason, U.; Gullberg, A.; Janke, A.
   #journal         Mol. Biol. Evol. (1997) 14:762-768
   #title           Phylogenetic analysis of mitochondrial DNA suggest a sister
                      group relationship between Xenarthra (Edentata) and
                      Ferungulates.
   #accession       I11441
      ##status          preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues        1-318 ##label ARN
      ##cross-references EMBL:Y11832, NID:e1044456; PID:e311016;
                        PIDN:CAA72529.1
GENETICS
   #gene            NADH1
   #genome          mitochondrion
   #genetic_code    SGC1
CLASSIFICATION      #superfamily NADH dehydrogenase (ubiquinone) chain 1
KEYWORDS            mitochondrion; NAD; oxidoreductase
SUMMARY             #length 318  #molecular-weight 35581  #checksum 5230

  Query Match             30.8%;  Score 77;  DB 2;  Length 318;
  Best Local Similarity   30.3%;  Pred. No. 3.32e+00;
  Matches    10;  Conservative    10;  Mismatches  12;  Indels    1;  Gaps    1;

Db     233 MMNALTAILFFGAMHNPMFPELHTLNLITKTLI 265
           |  |: : |:| :   : :: || | | |:::|:
Qy       1 MRPAIPTQLYFPT-DRSFLAILHQLTLTTMSLV 32


RESULT   10
ENTRY               A25330      #type complete
TITLE               bacteriochlorophyll a protein - Prosthecochloris aestuarii
ORGANISM            #formal_name Prosthecochloris aestuarii
DATE                30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change
                      18-Jun-1993
ACCESSIONS          A25330
REFERENCE           A25330
   #authors         Daurat-Larroque, S.T.; Brew, K.; Fenna, R.E.
   #journal         J. Biol. Chem. (1986) 261:3607-3615
   #title           The complete amino acid sequence of a bacteriochlorophyll
                      a-protein from Prosthecochloris aestuarii.
   #cross-references MUID:86140151
   #accession       A25330
      ##molecule_type protein
      ##residues        1-366 ##label DAU
SUMMARY             #length 366  #molecular-weight 40276  #checksum 1351

  Query Match             30.8%;  Score 77;  DB 2;  Length 366;
  Best Local Similarity   36.4%;  Pred. No. 3.32e+00;
  Matches    12;  Conservative     3;  Mismatches  17;  Indels    1;  Gaps    1;

Db     129 VRRNIPNPIYMQ-GRQFHDILMKVPLDNNDLVD 160
           :|  ||  :|    | |  ||  : |    |||
Qy       1 MRPAIPTQLYFPTDRSFLAILHQLTLTTMSLVD 33


RESULT   11
ENTRY               XHHU3       #type complete
TITLE               antithrombin III precursor - human
ORGANISM            #formal_name Homo sapiens #common_name man
DATE                05-Apr-1983 #sequence_revision 05-Apr-1983 #text_change
                      18-Jun-1999
ACCESSIONS          A49494; A93453; A93943; A92431; A33805; A94445; A34190;
                      S02530; S10716; A29371; I37191; I52399; I65277; I65278;
                      I65279; A44935; B44935; S63600; A01242; I37192; S46128
REFERENCE           A49494
   #authors         Olds, R.J.; Lane, D.A.; Chowdhury, V.; De Stefano, V.; Leone,
```

G.; Theim, S.L.
#journal Biochemistry (1993) 32:4216-4224
#title Complete nucleotide sequence of the antithrombin gene: evidence for homologous recombination causing thrombophilia.
#cross-references MUID:93237227
#accession A49494
##molecule_type DNA
##residues 1-464 ##label OLD
##cross-references EMBL:X68793; NID:g28906; PIDN:CAA48690.1; PID:g28907
##note sequence extracted from NCBI backbone (NCBIN:130247, NCBIP:130248)

REFERENCE A93453
#authors Bock, S.C.; Wion, K.L.; Vehar, G.A.; Lawn, R.M.
#journal Nucleic Acids Res. (1982) 10:8113-8125
#title Cloning and expression of the cDNA for human antithrombin III.
#cross-references MUID:83143280
#accession A93453
##molecule_type mRNA
##residues 1-464 ##label BOC1
##cross-references GB:L00190; GB:J00102; GB:J00103; GB:J00104; NID:g179128; PIDN:AAB40025.1; PID:g179130

REFERENCE A93943
#authors Chandra, T.; Stackhouse, R.; Kidd, V.J.; Woo, S.L.C.
#journal Proc. Natl. Acad. Sci. U.S.A. (1983) 80:1845-1848
#title Isolation and sequence characterization of a cDNA clone of human antithrombin III.
#cross-references MUID:83169777
#accession A93943
##molecule_type mRNA
##residues 1-464 ##label CHA
##cross-references GB:L00190; GB:J00102; GB:J00103; GB:J00104; NID:g179128; PIDN:AAB40025.1; PID:g179130

REFERENCE A92431
#authors Prochownik, E.V.; Markham, A.F.; Orkin, S.H.
#journal J. Biol. Chem. (1983) 258:8389-8394
#title Isolation of a cDNA clone for human antithrombin III.
#cross-references MUID:83238456
#accession A92431
##molecule_type mRNA
##residues 42-96,'R',98-464 ##label PRO
##note the authors translated the codon GAC for residue 206 as Asn, GGC for residue 298 as Glu, and TGC for residue 411 as Gln

REFERENCE A33305
#authors Chang, J.Y.
#journal J. Biol. Chem. (1989) 264:3111-3115
#title Binding of heparin to human antithrombin III activates selective chemical modification at lysine 236. Lys-107, Lys-125, and Lys-136 are situated within the heparin binding site of antithrombin III.
#cross-references MUID:89123426
#accession A33305
##molecule_type protein
##residues 90-105;124-143;147-161;165-171;268-273;381-402,446-457 ##label CH2

REFERENCE A94445
#authors Petersen, T.E.; Dudek-Wojciechowska, G.; Sottrup-Jensen, L., Magnusson, S.
#book in The Physiological Inhibitors of Blood Coagulation and Fibrinolysis, Collen, D., Wiman, B., and Verstraete, M., eds., pp.43-54, Elsevier/North Holland Biomedical Press, Amsterdam, 1979
#title Primary structure of antithrombin-III (heparin cofactor). Partial homology between alpha-1-antitrypsin and antithrombin-III.
#contents carbohydrate-binding site and disulfide bonds
#accession A94445
##molecule_type protein
##residues 33-68,'QE',71-242;244-245,248-250,254-464 ##label PET

REFERENCE A34190
#authors Zettlmeissl, G.; Conradt, H.S.; Nimtz, M.; Karges, H.E.
#journal J. Biol. Chem. (1989) 264:21153-21159
#title Characterization of recombinant human synthesized in Chinese hamster [illegible]
#cross-references MUID:[illegible]
#accession A34190
##molecule_type protein
##residues 33-39 ##label [illegible]
##experimental_source recombinant protein from [illegible] (CHO) cells

REFERENCE S02530
#authors Liu, C.S.; Chang, J.Y.
#journal Eur. J. Biochem. (1987) 167:247-[illegible]
#title Probing the heparin binding domain of [illegible] with V8 protease.
#cross-references MUID:[illegible]
#accession S02530
##status preliminary
##molecule_type protein
##residues [illegible] 189-202;210-214,'X';265-269,'X' ##label LIU

REFERENCE S10716
#authors Borg, J.Y.; Brennan, S.O.; Carrell, R. [illegible] D.J.; Shaw, J.
#journal FEBS Lett. (1990) 266:163-166
#title Antithrombin Pittsburgh IV 24 Arg>Cys. Its contribution to heparin binding.
#cross-references MUID:[illegible]
#accession S10716
##molecule_type protein
##residues 53-[illegible] ##label [illegible]
##note variant form Rouen-IV, 56-Cys, wa[illegible]

REFERENCE A29371
#authors Koide, T.; Odani, S.; Takahashi, K.; [illegible]
#journal Proc. Natl. Acad. Sci. U.S.A. (1984) [illegible]
#title Antithrombin III Toyama: replacement of [illegible] cysteine in hereditary abnormal antithrombin with [illegible] heparin-binding ability.
#cross-references MUID:[illegible]
#accession A29371
##molecule_type protein
##residues 53-[illegible],'G',73-79,'C',80-135 ##label [illegible]
##note variant Toyama with a substitution [illegible] consequently cannot bind heparin

REFERENCE I37191
#authors Bock, S.C.; Levitan, D.J.
#journal Nucleic Acids Res. (1983) 11:8569-8588
#title Characterization of an unusual DNA lesion in the human antithrombin III gene.
#cross-references MUID:[illegible]
#accession I37191
##status preliminary; translated from GB/EM[illegible]
##molecule_type DNA
##residues 1-14 ##label BOC2
##cross-references EMBL:X00237; NID:g28917; PIDN:[illegible] EMBL:X00238; NID:g28919; PID:g[illegible]

REFERENCE I52399
#authors Bock, S.C.; Marrinan, J.A.; Radziejewsk[illegible]
#journal Biochemistry (1988) 27:6171-6178
#title Antithrombin III Utah: proline-407 to [illegible] highly conserved region near the int[illegible]
#cross-references MUID:[illegible]
#note a published erratum appears in Biochem. [illegible]
#accession I52399
##status translation not shown; translated [illegible]
##molecule_type DNA
##residues 137-298 ##label BOC3
##cross-references GB:M21643; NID:g179147; PIDN:AA[illegible]
#accession I65277
##status translation not shown; translated [illegible]
##molecule_type DNA
##residues 137-254 ##label BOC4
##cross-references GB:M21643; NID:g179148; PIDN:AA[illegible]

```
   #accession     I65278
      ##status          translation not shown; translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues       407 464 ##label BOC5
      ##cross-references GB:M21645; NID:g179149; PIDN:AAA51795.1; PID:g179152
   #accession     I65279
      ##status          translation not shown; translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues       1-438,'L',440-464 ##label BOC6
      ##cross-references GB:M21642; NID:g179159; PIDN:AAA51796.1; PID:g179161
      ##note           mutant Utah
REFERENCE         A44935
   #authors       Grundy, C.B.; Thomas, P.; Millar, D.S.; Krawczak, M.;
                    Melissari, E.; Lindo, V.; Moffat, E.; Kakkar, V.V.; Cooper,
                    D.N.
   #journal       Blood (1991) 78:1027-1042
   #title         Recurrent deletion in the human antithrombin III gene.
   #cross-references MUID:91329813
   #accession     A44935
      ##molecule_type DNA
      ##residues       271-276,'CPVQHL' ##label CR1
      ##cross-references GB:S49757; NID:g233568; PIDN:AAB19467.1; PID:g233569
      ##experimental_source antithrombin III deficiency patient 1
      ##note          sequence extracted from NCBI backbone (NCBIN:49757,
                       NCBIP:49758)
   #accession     B44935
      ##molecule_type DNA
      ##residues       271-276,'VVFSIYDVPCRQVPLSARG' ##label CR2
      ##cross-references GB:S49759; NID:g233570; PIDN:AAB19468.1; PID:g233571
      ##experimental_source antithrombin III deficiency patient 2
      ##note          sequence extracted from NCBI backbone (NCBIN:49759,
                       NCBIP:49760)
      ##note          different frameshift mutations at an identical site in
                       unrelated patients suggests a mutational hot spot
REFERENCE         S63599
   #authors       Chang, W.S.W.; Wardell, M.R.; Lomas, D.A.; Carrell, R.W.
   #journal       Biochem. J. (1996) 314:647-653
   #title         Probing serpin reactive-loop conformations by proteolytic
                    cleavage.
   #cross-references MUID:96239126
   #accession     S63600
      ##status          preliminary
      ##molecule_type protein
      ##residues       414-428 ##label CB3
...
Note: remainder of annotations omitted.

  Query Match             30.8%;  Score 77;  DB 1;  Length 464;
  Best Local Similarity   24.2%;  Pred. No. 3.22e-02;
  Matches     8;  Conservative    15;  Mismatches   10;  Indels    0;  Gaps    0;

Db    425 RSLNPNRVTFKANRPFLVFIREVPLNTIIFMGR 457
          |:  | :: | ::|:|| ::::: | |: :: :
Qy      2 PPAIPTQLYFPTDRSFLAIIHQLTITTMSIVPQ 34


RESULT   12
ENTRY             S29133      #type complete
TITLE             urate oxidase (EC 1.7.3.3) - fruit fly (Drosophila virilis)
ALTERNATE_NAMES   uricase
ORGANISM          #formal_name Drosophila virilis
DATE              10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change
                    10-Sep-1999
ACCESSIONS        S29133
REFERENCE         S29133
   #authors       Wallrath, L.L.; Burnett, J.B.; Friedman, T.B.
   #submission    submitted to the EMBL Data Library, January 1991
   #accession     S29133
      ##molecule_type DNA
      ##residues       1-342 ##label WAT
      ##cross-references EMBL:X57134; NID:g9211; PIDN:CAA40397.1; PID:g9212
GENETICS
   #gene          FlyBase:Dvir/Uro
      ##cross-references FlyBase:FBgn0014101
   #status        162,1
CLASSIFICATION    #superfamily urate oxidase
KEYWORDS          copper binding; oxidoreductase; peroxisome
FEATURE
   340 342             #region peroxisome/glyoxysome location signal
                         (S-[EFH] ) motif\
   160,162             #binding_site copper (His) (type 2) #status predicted
SUMMARY           #length 342  #molecular-weight 38724  #checksum 2138

  Query Match             30.4%;  Score 76;  DB 1;  Length 342;
  Best Local Similarity   24.1%;  Pred. No. 4.57e+00;
  Matches     7;  Conservative    13;  Mismatches   8;  Indels    1;  Gaps    1;

Db    287 PSVQHTLYL-SEKQVLDVLPQVSVVSMTM 294
          |..  ||. :::  | :| |::: :|::
Qy      8 PAIPTQLYFPTDRSFLAILHQLTITTMSL 31


RESULT   13
ENTRY             A53077      #type complete
TITLE             ubiquinol- cytochrome c reductase (EC 1.10.2.2) cytochrome b
                    Spermophilus richardsonii mitochondrion
ORGANISM          #formal_name mitochondrion Spermophilus richardsonii
DATE              04-Sep-1997 #sequence_revision 04-Sep-1997 #text_change
                    26-Aug-1999
ACCESSIONS        A53077
REFERENCE         A53077
   #authors       Thomas, W.K.; Martin, S.L.
   #journal       Mol. Phylogenet. Evol. (1993) 2:330-336
   #title         A recent origin of marmots.
   #cross-references MUID:94326000
   #accession     A53077
      ##status          preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues       1-379 ##label RES
      ##cross-references GB:S73150; NID:g639944; PIDN:AAC60686.1; PID:g639945
GENETICS
   #genome        mitochondrion
   #genetic_code SGC1
CLASSIFICATION    #superfamily cytochrome b, cytochrome b homology, cytochrome
                    b6 homology; plastoquinol--plastocyanin reductase 17K
                    protein homology
KEYWORDS          chromoprotein; electron transfer; heme; iron; mitochondrion;
                    oxidative phosphorylation; oxidoreductase; respiratory
                    chain; transmembrane protein
FEATURE
   11-339              #domain cytochrome b homology #label CBH\
   11 339              #domain cytochrome b6 homology #label CB6\
   221-339             #domain plastoquinol--plastocyanin reductase 17K protein
                         homology #label 17K\
   83,196              #binding_site heme iron, low potential (His) (axial
                         ligands) #status predicted\
   97,201              #binding_site heme iron, high potential (His) (axial
                         ligands) #status predicted
SUMMARY           #length 379  #molecular weight 42924  #checksum 6769

  Query Match             30.0%;  Score 75;  DB 2;  Length 379;
  Best Local Similarity   39.3%;  Pred. No. 6.26e+00;
  Matches    11;  Conservative    8;  Mismatches   7;  Indels    2;  Gaps    2;

Db    210 IPFHPYY TIKDILGVL LLILALMTLV 243
          || : |: | : :|::|  | |: |:||
Qy      5 IPIQLYFPTDRSFLAILHQLTLLTIMSLV 42


RESULT   14
ENTRY             D71542      #type complete
TITLE             probable ribose-5-phosphate isomerase A - Chlamydia
                    trachomatis (serotype D, strain UW3/Cx)
ORGANISM          #formal_name Chlamydia trachomatis
DATE              13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change
                    26-Aug-1999
```

```
ACCESSIONS        D71542
REFERENCE         A71570
   #authors       Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe,
                    R.; Aravind, L.; Mitchell, W.P.; Olinger, L.; Tatusov,
                    R.L.; Zhao, Q.; Koonin, E.V.; Davis, R.W.
   #journal       Science (1998) 282:754-759
   #title         Genome sequence of an obligate intracellular pathogen of
                    humans: Chlamydia trachomatis.
   #cross-references MUID:99000809
   #accession     D71542
      ##status          preliminary
      ##molecule_type DNA
      ##residues       1-242 ##label ARN
      ##cross-references GB:AE001295; GB:AE001273; NID:g3328617; PID:g3328621
      ##experimental_source serotype D, strain UW-3/Cx
GENETICS
   #gene          rpiA
CLASSIFICATION    #superfamily Haemophilus influenzae ribose-5-phosphate
                    isomerase
SUMMARY           #length 242  #molecular-weight 26646  #checksum 8729

  Query Match             29.6%;  Score 74;  DB 2;  Length 242;
  Best Local Similarity   44.8%;  Pred. No. 8.55e+00;
  Matches     8;  Conservative    11;  Mismatches    3;  Indels    1;  Gaps    1;

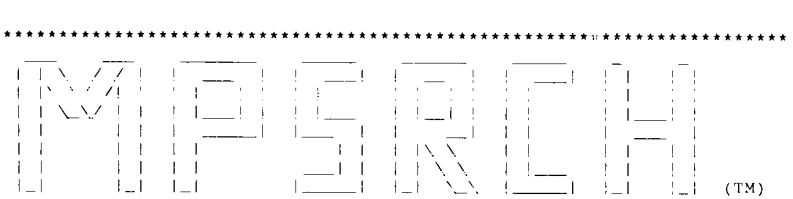
Db       9 FSSDK-FFPIKQKLALEAVALVE 30
           |:::|: |:::| : |:| ::::||:
Qy      11 FPTDRSFLAILHQLTLLTIMSLVD 33


RESULT   15
ENTRY             S68362      #type fragment
TITLE             glucose transport protein type 2 - rat (fragment)
ORGANISM          #formal_name Rattus norvegicus #common_name Norway rat
DATE              15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change
                    07-May-1999
ACCESSIONS        S68362
REFERENCE         S68362
   #authors       Ahn, Y.; Kim, J.; Han, G.; Lee, B.; Kim, Y.
   #journal       Arch. Biochem. Biophys. (1995) 323:387-396
   #title         Cloning and characterization of rat pancreatic
                    beta-cell/liver type glucose transporter gene: a unique
                    exon/intron organization.
   #cross-references MUID:96063615
   #accession     S68362
      ##status          preliminary; nucleic acid sequence not shown;
                         translation not shown
      ##molecule_type DNA
      ##residues       1-522 ##label AHN
      ##cross-references EMBL:L28134
      ##note           the nucleotide sequence was submitted to the EMBL Data
                         Library, September 1994
GENETICS
   #introns       5/1  [illegible]  456/3
CLASSIFICATION    #superfamily glucose transport protein
SUMMARY           #length 522  #checksum 4585

  Query Match             29.6%;  Score 74;  DB 2;  Length 522;
  Best Local Similarity   50.0%;  Pred. No. 8.55e+00;
  Matches    12;  Conservative     4;  Mismatches    7;  Indels    1;  Gaps    1;

Db     179 PTILRGAIGTLHQLALVIGILISQ 202
           ||  |: |: |||||:| |  |: :
Qy      12 PTD RSFLAILHQLILITMSLVDQ 34


Search completed: Fri Aug 25 18:04:28 2000
Job time : 32 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 25 18:06:30 2000; MasPar time 2.37 Seconds
219.705 Million cell updates/sec

Tabular output not generated.

Title: >US-09-257-585-7
Description: (1-34) from US09257585.pep
Perfect Score: 250
Sequence: 1 MRPAIPTQLYFPTDRSFLAILHQLTLITMSLVDQ 34

Scoring table: PAM 150
Gap 11

Searched: 152433 seqs, 15329240 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:6_COMB 4:PCT_COMB 5:backfiles1

Statistics: Mean 22.494; Variance 94.966; scale 0.237

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	82	32.8	289	4	PCT-US92-0	Sequence 15, Applicati	6.59e+00
2	80	32.0	327	2	US-08-562	Sequence 12, Applicati	9.57e+00
3	80	32.0	327	1	US-08-475-	Sequence 12, Applicati	9.57e+00
4	80	32.0	465	3	US-08-948-	Sequence 6, Applicatio	9.57e+00
5	77	30.8	464	1	US-08-046-	Sequence 2, Applicatio	1.67e+01
6	73	29.2	34	1	US-08-781-	Sequence 4, Applicatio	3.46e+01
7	69	27.6	534	2	US-09-031-	Sequence 4, Applicatio	7.12e+01
8	67	26.8	525	1	US-08-461-	Sequence 17, Applicati	1.02e+02
9	67	26.8	525	1	US-08-461	Sequence 15, Applicati	1.02e+02
10	67	26.8	525	1	US-08-077	Sequence 15, Applicati	1.02e+02
11	67	26.8	525	1	US-08-465-	Sequence 17, Applicati	1.02e+02
12	67	26.8	525	1	US-08-461	Sequence 17, Applicati	1.02e+02
13	67	26.8	525	1	US-08-465-	Sequence 15, Applicati	1.02e+02
14	67	26.8	525	1	US-08-461-	Sequence 15, Applicati	1.02e+02
15	67	26.8	525	1	US-08-077-	Sequence 17, Applicati	1.02e+02
16	66	26.4	358	2	US-08-463-	Sequence 6, Applicatio	1.21e+02
17	66	26.4	358	2	US-08-462-	Sequence 6, Applicatio	1.21e+02
18	66	26.4	358	3	US-08-652-	Sequence 6, Applicatio	1.21e+02
19	66	26.4	358	2	US-08-461-	Sequence 6, Applicatio	1.21e+02
20	66	26.4	358	3	US-08-463	Sequence 6, Applicatio	1.21e+02
21	66	26.4	358	3	US-08-465	Sequence 6, Applicatio	1.21e+02
22	66	26.4	383	2	US-09-031	Sequence 3, Applicatio	1.21e+02
23	66	26.4	433	2	US-08-480	Sequence 7, Applicatio	1.21e+02
24	66	26.4	433	1	US-08-488-	Sequence 7, Applicatio	1.21e+02
25	66	26.4	433	1	US-08-522	Sequence 7, Applicatio	1.21e+02
26	66	26.4	661	4	PCT-US91-0	Sequence 2, Applicatio	1.21e+02
27	66	26.4	1079	2	US-08-929-	Sequence 8, Applicatio	1.21e+02
28	65	26.0	467	3	US-08-179-	Sequence 4, Applicatio	1.44e+02
29	65	26.0	516	1	US-08-638-	Sequence 4, Applicatio	1.44e+02
30	65	26.0	1040	2	US-08-254-	Sequence 2, Applicatio	1.44e+02
31	65	26.0	1074	1	US-07-912-	Sequence 4, Applicatio	1.44e+02
32	65	26.0	2873	2	US-08-464-	Sequence 1, Applicatio	1.44e+02
33	65	26.0	2873	4	PCT-US95-0	Sequence 1, Applicatio	1.44e+02
34	65	26.0	2873	2	US-08-444-	Sequence 1, Applicatio	1.44e+02
35	65	26.0	2873	1	US-08-638-	Sequence 2, Applicatio	1.44e+02
36	65	26.0	2873	2	US-08-461-	Sequence 1, Applicatio	1.44e+02
37	65	26.0	2873	1	US-08-466	Sequence 1, Applicatio	1.44e+02
38	65	26.0	2873	2	US-08-485-	Sequence 1, Applicatio	1.44e+02
39	65	26.0	2873	2	US-08-461	Sequence 1, Applicatio	1.44e+02
40	65	26.0	2910	4	PCT-US95-0	Sequence 1, Applicatio	1.44e+02
41	65	26.0	2910	2	US-08-444	Sequence 1, Applicatio	1.44e+02
42	65	26.0	2910	1	US-08-466-	Sequence 1, Applicatio	1.44e+02
43	65	26.0	2910	2	US-08-464	Sequence 1, Applicatio	1.44e+02
44	65	26.0	2910	2	US-08-485-	Sequence 1, Applicatio	1.44e+02
45	64	25.6	1285	2	US-08-319-	Sequence 1, Applicatio	1.71e+02

ALIGNMENTS

RESULT 1
ID PCT-US92-00282-15 STANDARD; PRT; 289 AA
XX
AC xxxxxx
XX
DT
XX
DE Sequence 15, Application PC/US92200282
XX
CC Sequence 15, Application PC/US9200282
CC GENERAL INFORMATION:
CC APPLICANT: OWENS, U.A.S.
CC APPLICANT: RITTER, J.SEPH K.
CC TITLE OF INVENTION: THE GENETIC LOCUS [illegible] AS
CC TITLE OF INVENTION: THEREIN.
CC NUMBER OF SEQUENCES: 4
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: CUSHMAN DARBY & CUSHMAN
CC STREET: 1615 L STREET, N.W.
CC CITY: WASHINGTON
CC STATE: D.C.
CC COUNTRY: U.S.A.
CC ZIP: 20036-5601
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC Compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PC/US92200282
CC FILING DATE: 1992[illegible]
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: SCOTT, WATSON I.
CC REGISTRATION NUMBER: 2[illegible]
CC REFERENCE/DOCKET NUMBER: 91542 PCT
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202 861-[illegible]
CC TELEFAX: 202 822-0944
CC TELEX: 6714627 CUSH
CC INFORMATION FOR SEQ ID NO: 15:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 289 amino acids
CC TYPE: AMINO ACID
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein

```
SQ   SEQUENCE   289 AA; 32971 MW; 450461 CN;

  Query Match           32.8%;  Score 82;  DB 4;  Length 289;
  Best Local Similarity  46.4%;  Pred. No. 6.59e+00;
  Matches   13;  Conservative    7;  Mismatches   5;  Indels   3;  Gaps   3;

Db      97 TQLYFETEH-FLKKFFRSMAMLNNMSLV 123
           ||||| ::: ||  ::: :: |  ||||
Qy       7 TQLYFPTDRSFLA-ILHQLF-LTTMSLV 32


RESULT    2
ID   US-08-562-114B-12     STANDARD;     PRT;   327 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 12, Application US/08562114B
XX
CC    Sequence 12, Application US/08562114B
CC    Patent No. 5972646
CC     GENERAL INFORMATION:
CC       APPLICANT:  ERIKSSON ET AL.
CC       TITLE OF INVENTION:  ISOLATED NUCLEIC ACID MOLECULE WHICH CODES FOR A
CC       TITLE OF INVENTION:  32 KDA PROTEIN HAVING 11-CIS RETINOL HYROGENASE
CC       TITLE OF INVENTION:  ACTIVITY, AND WHICH ASSOCIATES WITH P63, A
CC       TITLE OF INVENTION:  PORTION OF A RETINOL BINDING PROTEIN RECEPTOR
CC       NUMBER OF SEQUENCES:  31
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Felfe & Lynch
CC         STREET:  805 Third Avenue
CC         CITY:  New York City
CC         STATE:  New York
CC         ZIP:  10022
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette, 3.5 inch, 144 kb storage
CC         COMPUTER:  IBM
CC         OPERATING SYSTEM:  PC-DOS
CC         SOFTWARE:  Wordperfect 5.1 and ASCII
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/562,114B
CC         FILING DATE:  22-No. 5972646ember 1995
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  08/375,962
CC         FILING DATE:  20-January-1995
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Kohli, Vineet
CC         REGISTRATION NUMBER:  37,003
CC         REFERENCE/DOCKET NUMBER:  LUD 5372.2 CIP
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (212) 698 9200
CC         TELEFAX:  (212) 838 3884
CC     INFORMATION FOR SEQ ID NO:  12:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  327 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
CC       FEATURE:
CC         NAME/KEY:  Human estradiol 17-b dehydrogenase (HEDH)
SQ   SEQUENCE   327 AA; 34849 MW; 520571 CN;

  Query Match           32.0%;  Score 80;  DB 2;  Length 327;
  Best Local Similarity  45.5%;  Pred. No. 9.57e+00;
  Matches   10;  Conservative    7;  Mismatches   4;  Indels   1;  Gaps   1;

Db     244 LRAPKPTLRYFTTER-FLPLLR 264
           :|:: ||  || |:| ||::|:
Qy       1 MRPAIPTQLYFPTDRSFLAILH 22


RESULT    3
ID   US-08-375-962B-12     STANDARD;     PRT;   327 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 12, Application US/08375962B
XX
CC    Sequence 12, Application US/08375962B
CC    Patent No. 5731195
CC     GENERAL INFORMATION:
CC       APPLICANT:  SIMON, ANDRAS; HELLMAN, ULF; WERNSTEDT,
CC       APPLICANT:  CHRISTER; ERIKSSON, ULF.
CC       TITLE OF INVENTION:  Isolated Nucleic Acid Molecule
CC       TITLE OF INVENTION:  Which Codes for A 32 kDa Protein Having 11-CIS Retinol
CC       TITLE OF INVENTION:  Dehydrogenase Activity, and Which Associates With P63,
CC       TITLE OF INVENTION:  a Portion of a Retinol Binding Protein Receptor
CC       NUMBER OF SEQUENCES:  14
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Felfe & Lynch
CC         STREET:  805 Third Avenue
CC         CITY:  New York City
CC         STATE:  New York
CC         ZIP:  10022
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette, 3.5 inch, 144 kb storage
CC         COMPUTER:  IBM
CC         OPERATING SYSTEM:  PC-DOS
CC         SOFTWARE:  Wordperfect (ASCII standard)
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/375,962B
CC         FILING DATE:  20-January 1995
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  08/258,418
CC         FILING DATE:  6-October-1994
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Pasqualini, Patricia A.
CC         REGISTRATION NUMBER:  34,894
CC         REFERENCE/DOCKET NUMBER:  LUD 5372.1 CIP
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (212) 688-9200
CC         TELEFAX:  (212) 838-3884
CC     INFORMATION FOR SEQ ID NO:  12:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  327 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
CC       FEATURE:
CC         NAME/KEY:  Human estradiol 17 b dehydrogenase (hEDH)
SQ   SEQUENCE   327 AA; 34849 MW; 520571 CN;

  Query Match           32.0%;  Score 80;  DB 1;  Length 327;
  Best Local Similarity  45.5%;  Pred. No. 9.57e+00;
  Matches   10;  Conservative    7;  Mismatches   4;  Indels   1;  Gaps   1;

Db     244 LRAPKPTLRYFTTER-FLPLLR 264
           :|:: ||  || |:| ||::|:
Qy       1 MRPAIPTQLYFPTDRSFLAILH 22


RESULT    4
ID   US-08-948-997-6     STANDARD;     PRT;   465 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 6, Application US/08948997
XX
```

```
CC    Sequence 6, Application US/08948997
CC    Patent No. 6008020
CC     GENERAL INFORMATION:
CC       APPLICANT:  HASTINGS, GREGG
CC       APPLICANT:  COLEMAN, TIM
CC       APPLICANT:  LAWRENCE, DANIEL
CC       TITLE OF INVENTION:  BRAIN-ASSOCIATED INHIBITOR OF
CC       TITLE OF INVENTION:  TISSUE-TYPE PLASMINOGEN ACTIVATOR
CC       NUMBER OF SEQUENCES:  17
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  HUMAN GENOME SCIENCES, INC.
CC         STREET:  9410 KEY WEST AVENUE
CC         CITY:  ROCKVILLE
CC         STATE:  MD
CC         COUNTRY:  USA
CC         ZIP:  20850
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/948,997
CC         FILING DATE:  Oct-10-97
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  A. ANDERS BROOKES
CC         REGISTRATION NUMBER:  36,373
CC         REFERENCE/DOCKET NUMBER:  PF336
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (301) 309-8504
CC         TELEFAX:  (301) 309-8512
CC     INFORMATION FOR SEQ ID NO:  6:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  465 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   465 AA; 52003 MW; 1130566 CN;

  Query Match           32.0%;  Score 80;  DB 3;  Length 465;
  Best Local Similarity  24.2%;  Pred. No. 9.57e+00;
  Matches    8;  Conservative   16;  Mismatches   9;  Indels   0;  Gaps   0;

Db    426 RSLNPNRVTFKANRPFLVLIRFVALNTIIFMGR 458
          |: | :: | ::|:|| ::::::: |: :: :
Qy      2 RPAIPTQLYFPTDRSFIAIHQITIIIMSIVDQ 34


RESULT    5
ID   US-08-046-431A-2      STANDARD;      PRT;   464 AA.
XX
AC   XXXXXX
XX
DT
XX
DE   Sequence 2, Application US/08046431A
XX
CC    Sequence 2, Application US/08046431A
CC    Patent No. 5420252
CC     GENERAL INFORMATION:
CC       APPLICANT:  KATO, HIROYUKI
CC       APPLICANT:  :  YOSHIIAKE, SHINJI
CC       APPLICANT:  :  SUZUKI, SUGURA
CC       APPLICANT:  :  SUZUKI, NOBORU
CC       APPLICANT:  :  SETO, TOSHIO
CC       APPLICANT:  :  NAGAOKA, NAOKO
CC       APPLICANT:  :  MIZUI, YOSHIHARU
CC       TITLE OF INVENTION:  HUMAN ANTITHROMBIN III MUTANT
CC       NUMBER OF SEQUENCES:  4
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  FLYNN, THIEL, BOUTELL & TANIS, P.C.
CC         STREET:  2026 Rambling Road
CC         CITY:  Kalamazoo
CC         STATE:  Michigan
CC         COUNTRY:  USA
CC         ZIP:  49008-1699
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette, 5.5 inches, 1.44 MB
CC         COMPUTER:  IBM PC/XT/AT Compatible
CC         OPERATING SYSTEM:  MS DOS 5.0
CC         SOFTWARE:  WordPerfect 5.0
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/046,431A
CC         FILING DATE:
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  JP 00488
CC         FILING DATE:  14-APR-1992
CC         APPLICATION NUMBER:  JP 41855
CC         FILING DATE:  22-FEB-1993
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Terryence F. Chapman
CC         REGISTRATION NUMBER:  32549
CC         REFERENCE/DOCKET NUMBER:  Furuya Case 1288
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (616) 381-1150
CC         TELEFAX:  (616) 381-5465
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  464 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:
CC         DESCRIPTION:  protein
SQ   SEQUENCE   464 AA; 52602 MW; 1117642 CN;

  Query Match           32.0%;  Score 77;  DB 1;  Length
  Best Local Similarity  24.2%;  Pred. No. 1.67e+01;
  Matches    8;  Conservative   15;  Mismatches   19;

Db    425 RSLNPNRVTFKANRPFLVLIRFVALNTIIFMGR 457
          |: | :: | ::|:| ::::: : |: :: :
Qy      2 RPAIPTQLYFPTDRSFIAILHQITIIIMSIVDQ 34


RESULT    6
ID   US-08-781-020-4      STANDARD;      PRT;   39 AA.
XX
AC   XXXXXX
XX
DT
XX
DE   Sequence 4, Application US/08781020
XX
CC    Sequence 4, Application US/08781020
CC    Patent No. 5792749
CC     GENERAL INFORMATION:
CC       APPLICANT:  Wright, H. Leslie
CC       TITLE OF INVENTION:  METHOD AND COMPOSITION FOR
CC       TITLE OF INVENTION:  DENSITY LIPOPROTEIN CHOLES
CC       NUMBER OF SEQUENCES:  15
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  WHITHAM, CURTIS, WHITHAM & McGINN
CC         STREET:  Reston Int'l. Center, 11800 Sunrise
CC         STREET:  Suite 900
CC         CITY:  Reston
CC         STATE:  VA
CC         COUNTRY:  USA
CC         ZIP:  20191
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1
```

```
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/09/781,020
CC         FILING DATE:  09-JAN-1997
CC         CLASSIFICATION:  514
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Whitham, Michael E.
CC         REGISTRATION NUMBER:  32,635
CC         REFERENCE/DOCKET NUMBER:  VCU-16
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  703-491-2510
CC         TELEFAX:  703-491-9035
CC     INFORMATION FOR SEQ ID NO:  4:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  39 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  peptide
SQ   SEQUENCE   39 AA; 4473 MW; 8270 CN;

  Query Match           29.2%;  Score 73;  DB 1;  Length 39;
  Best Local Similarity  24.1%;  Pred. No. 3.46e+01;
  Matches    7;  Conservative   14;  Mismatches   8;  Indels   0;  Gaps   0;

Db       4 PNRVTFKANRPFLVFIREVPLNTIIFMGR 32
           | :: | ::|:|| ::::: | |: :: :
Qy       6 PTQLYFPTDRSFLAILHQLTLIIMSLVDQ 34


RESULT    7
ID   US 09-041-492-4      STANDARD;      PRT;   534 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 4, Application US/09031392
XX
CC    Sequence 4, Application US/09031392
CC    Patent No. 5942398
CC     GENERAL INFORMATION:
CC       APPLICANT:  Tartaglia, Louis A.
CC       APPLICANT:  Weng, Xun
CC       TITLE OF INVENTION:  NUCLEIC ACID MOLECULES
CC       TITLE OF INVENTION:  ENCODING GLUTEX AND USES THEREOF
CC       NUMBER OF SEQUENCES:  10
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Fish & Richardson P.C.
CC         STREET:  225 Franklin Street
CC         CITY:  Boston
CC         STATE:  MA
CC         COUNTRY:  USA
CC         ZIP:  02110-2804
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette
CC         COMPUTER:  IBM Compatible
CC         OPERATING SYSTEM:  Windows95
CC         SOFTWARE:  FastSEQ for Windows Version 2.0
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/09/031,392
CC         FILING DATE:  26-FEB-1998
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Meiklejohn, Ph.D., Anita L.
CC         REGISTRATION NUMBER:  35,283
CC         REFERENCE/DOCKET NUMBER:  07334/072001
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  617/542-5070
CC         TELEFAX:  617/542-8906
CC         TELEX:  200154
CC     INFORMATION FOR SEQ ID NO:  4:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  534 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   534 AA; 57862 MW; 1553071 CN;

  Query Match           27.6%;  Score 69;  DB 2;  Length 534;
  Best Local Similarity  45.8%;  Pred. No. 7.12e+01;
  Matches   11;  Conservative    5,  Mismatches   7,  Indels   1,  Gaps   1;

Db     193 PTALRGALGTLHQLAIVTGILISQ 216
           ||  |: |: |||:: |  |: |
Qy      12 PTD-RSFLAILHQLTTMSLVDQ 34


RESULT    8
ID   US-08-461-599-17      STANDARD;      PRT;   525 AA.
XX
AC   vvvvvv
XX
DT
XX
DE   Sequence 17, Application US/08461599
XX
CC    Sequence 17, Application US/08461599
CC    Patent No. 5658777
CC     GENERAL INFORMATION:
CC       APPLICANT:  FUJII, Toshio
CC       APPLICANT:  IWAMATSU, Akihiro
CC       APPLICANT:  YOSHIMOTO, Hiroyuki
CC       APPLICANT:  MINETOKI, Toshitaka
CC       APPLICANT:  BOGAKI, Takayuki
CC       APPLICANT:  NAGASAWA, Naoshi
CC       TITLE OF INVENTION:  ALCOHOL ACETYLTRANSFERASE GENES AND USE
CC       TITLE OF INVENTION:  THEREOF
CC       NUMBER OF SEQUENCES:  19
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Foley & Lardner
CC         STREET:  3000 K Street, N.W., Suite 500
CC         CITY:  Washington, D.C.
CC         COUNTRY:  USA
CC         ZIP:  20007-5109
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/461,599
CC         FILING DATE:  05-JUN-1995
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 08/077,939
CC         FILING DATE:  18-JUN-1993
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  JP 62997/1993
CC         FILING DATE:  26-FEB-1993
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  JP 184328/1992
CC         FILING DATE:  18-JUN-1992
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  BENI, Stephen A.
CC         REGISTRATION NUMBER:  29,768
CC         REFERENCE/DOCKET NUMBER:  49441/105 EYPA
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (202)672-5300
CC         TELEFAX:  (202)672-5399
CC         TELEX:  904136
CC     INFORMATION FOR SEQ ID NO:  17:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  525 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   525 AA; 61064 MW; 1459439 CN;
```

```
  Query Match           26.8%;  Score 67;  DB 1;  Length 525;
  Best Local Similarity  25.0%;  Pred. No. 1.02e+02;
  Matches    5;  Conservative   10;  Mismatches   5;  Indels   0;  Gaps   0;

Db     106 HDYISVLQELKLSGVVLNEQ 125
           : :::::|:| |: : | :|
Qy      15 RSFLAILHQLILITMSIVDQ 34


RESULT    9
ID   US-08-461-599-15      STANDARD;      PRT:   525 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 15, Application US/08461599
XX
CC    Sequence 15, Application US/08461599
CC    Patent No. 5658777
CC     GENERAL INFORMATION:
CC       APPLICANT:  FUJII, Toshio
CC       APPLICANT:  IWAMATSU, Akihiro
CC       APPLICANT:  YOSHIMOTO, Hiroyuki
CC       APPLICANT:  MINETOKI, Toshitaka
CC       APPLICANT:  BOGAKI, Takayuki
CC       APPLICANT:  NAGASAWA, Naoshi
CC       TITLE OF INVENTION:  ALCOHOL ACETYLTRANSFERASE GENES AND USE
CC       TITLE OF INVENTION:  THEREOF
CC       NUMBER OF SEQUENCES:  19
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Foley & Lardner
CC         STREET:  3000 K Street, N.W., Suite 500
CC         CITY:  Washington, D.C.
CC         COUNTRY:  USA
CC         ZIP:  20007-5109
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  Patentin Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/461,599
CC         FILING DATE:  05-JUN-1995
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 08/077,939
CC         FILING DATE:  18-JUN-1993
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  JP 62997/1993
CC         FILING DATE:  26-FEB-1993
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  JP 184328/1992
CC         FILING DATE:  18-JUN-1992
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  BENT, Stephen A.
CC         REGISTRATION NUMBER:  29,768
CC         REFERENCE/DOCKET NUMBER:  34441/185 KYPA
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (202)672-5300
CC         TELEFAX:  (202)672-5399
CC         TELEX:  904136
CC     INFORMATION FOR SEQ ID NO:  15:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  525 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   525 AA; 61062 MW; 1464200 CN;

  Query Match           26.8%;  Score 67;  DB 1;  Length 525;
  Best Local Similarity  25.0%;  Pred. No. 1.02e+02;
  Matches    5;  Conservative   10;  Mismatches   5;  Indels   0;  Gaps   0;

Db     106 HDYISVLQELKLSGVVLNEQ 125
           : :::::|:| |: : | :|
Qy      15 RSFLAILHQLILITMSIVDQ 34


RESULT    10
ID   US-08-077-939-15      STANDARD;      PRT:   525 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 15, Application US/08077939
XX
CC    Sequence 15, Application US/08077939
CC    Patent No. 5521088
CC     GENERAL INFORMATION:
CC       APPLICANT:  FUJII, Toshio
CC       APPLICANT:  IWAMATSU, Akihiro
CC       APPLICANT:  YOSHIMOTO, Hiroyuki
CC       APPLICANT:  MINETOKI, Toshitaka
CC       APPLICANT:  BOGAKI, Takayuki
CC       APPLICANT:  NAGASAWA, Naoshi
CC       TITLE OF INVENTION:  ALCOHOL ACETYLTRANSFERASE
CC       TITLE OF INVENTION:  THEREOF
CC       NUMBER OF SEQUENCES:  19
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Foley & Lardner
CC         STREET:  3000 K Street, N.W., Suite 500
CC         CITY:  Washington, D.C.
CC         COUNTRY:  USA
CC         ZIP:  20007-5109
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  Patentin Release #1.0, Version #1
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/077,939
CC         FILING DATE:  18-JUN-1993
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  JP 184328/1992
CC         FILING DATE:  18-JUN-1992
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  JP 62997/1993
CC         FILING DATE:  26-FEB-1993
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  BENT, Stephen A.
CC         REGISTRATION NUMBER:  29,768
CC         REFERENCE/DOCKET NUMBER:  34441/185 KYPA
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (202)672-5300
CC         TELEFAX:  (202)672-5399
CC         TELEX:  904136
CC     INFORMATION FOR SEQ ID NO:  15:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  525 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   525 AA; 61062 MW; 1464200 CN;

  Query Match           26.8%;  Score 67;  DB 1;  Length 525;
  Best Local Similarity  25.0%;  Pred. No. 1.02e+02;
  Matches    5;  Conservative   10;  Mismatches   5;  Indels   0;  Gaps   0;

Db     106 HDYISVLQELKLSGVVLNEQ 125
           : :::::|:| |: : | :|
Qy      15 RSFLAILHQLILITMSIVDQ 34
```

```
RESULT   11
ID   US-08-465-334-17      STANDARD;      PRT;   525 AA.
XX
AC   xxxxxx
XX
DT
XX
LE   Sequence 17, Application US/08465334
XX
CC    Sequence 17, Application US/08465334
CC    Patent No. 5728412
CC     GENERAL INFORMATION:
CC       APPLICANT:  FUJII, Toshio
CC       APPLICANT:  IWAMATSU, Akihiro
CC       APPLICANT:  YOSHIMOTO, Hiroyuki
CC       APPLICANT:  MINETOKI, Toshitaka
CC       APPLICANT:  BOGAKI, Takayuki
CC       APPLICANT:  NAGASAWA, Naoshi
CC       TITLE OF INVENTION:  ALCOHOL ACETYLTRANSFERASE GENES AND USE
CC       TITLE OF INVENTION:  THEREOF
CC       NUMBER OF SEQUENCES:  19
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Foley & Lardner
CC         STREET:  3000 K Street, N.W., Suite 500
CC         CITY:  Washington, D.C.
CC         COUNTRY:  USA
CC         ZIP:  20007-5109
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/465,334
CC         FILING DATE:  05-JUN-1995
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 08/077,939
CC         FILING DATE:  18-JUN-1993
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  JP 62997/1993
CC         FILING DATE:  26-FEB-1993
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  JP 184328/1992
CC         FILING DATE:  18-JUN-1992
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  BENT, Stephen A.
CC         REGISTRATION NUMBER:  29,768
CC         REFERENCE/DOCKET NUMBER:  49441/101 KYPA
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (202)672-5300
CC         TELEFAX:  (202)672-5399
CC         TELEX:  904136
CC     INFORMATION FOR SEQ ID NO:  17:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  525 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   525 AA; 61064 MW; 1459439 CN;

  Query Match           26.8%;  Score 67;  DB 1;  Length 525;
  Best Local Similarity  25.0%;  Pred. No. 1.02e+02;
  Matches     5;  Conservative   10;  Mismatches   5;  Indels   0;  Gaps   0;

Db     106 HDYISVLQELKLSGVVLNEQ 125
           : ::::|:::| |: : | :|
Qy      15 RSFLAILHQLTLTTMSLVDQ 34


RESULT   12
ID   US-08-461-621-17      STANDARD;      PRT;   525 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 17, Application US/08461621
XX
CC    Sequence 17, Application US/08461621
CC    Patent No. 5686284
CC     GENERAL INFORMATION:
CC       APPLICANT:  FUJII, Toshio
CC       APPLICANT:  IWAMATSU, Akihiro
CC       APPLICANT:  YOSHIMOTO, Hiroyuki
CC       APPLICANT:  MINETOKI, Toshitaka
CC       APPLICANT:  BOGAKI, Takayuki
CC       APPLICANT:  NAGASAWA, Naoshi
CC       TITLE OF INVENTION:  ALCOHOL ACETYLTRANSFERASE GENES AND USE
CC       TITLE OF INVENTION:  THEREOF
CC       NUMBER OF SEQUENCES:  19
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Foley & Lardner
CC         STREET:  3000 K Street, N.W., Suite 500
CC         CITY:  Washington, D.C.
CC         COUNTRY:  USA
CC         ZIP:  20007-5109
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/461,621
CC         FILING DATE:  05-JUN-1995
CC         CLASSIFICATION:  424
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 08/077,939
CC         FILING DATE:  18-JUN-1993
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  JP 62997/1993
CC         FILING DATE:  26-FEB-1993
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  JP 184328/1992
CC         FILING DATE:  18-JUN-1992
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  BENT, Stephen A.
CC         REGISTRATION NUMBER:  29,768
CC         REFERENCE/DOCKET NUMBER:  49441/103 KYPA
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (202)672-5300
CC         TELEFAX:  (202)672-5399
CC         TELEX:  904136
CC     INFORMATION FOR SEQ ID NO:  17:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  525 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   525 AA; 61064 MW; 1459439 CN;

  Query Match           26.8%;  Score 67;  DB 1;  Length 525;
  Best Local Similarity  25.0%;  Pred. No. 1.02e+02;
  Matches     5;  Conservative   10;  Mismatches   5;  Indels   0;  Gaps   0;

Db     106 HDYISVLQELKLSGVVLNEQ 125
           : ::::|:::| |: : | :|
Qy      15 RSFLAILHQLTLTTMSLVDQ 34


RESULT   13
ID   US-08-465-334-15      STANDARD;      PRT;   525 AA.
XX
AC   xxxxxx
XX
```

```
DT
XX
DE   Sequence 15, Application US/08465334
XX
CC    Sequence 15, Application US/08465334
CC    Patent No. 5728812
CC     GENERAL INFORMATION:
CC       APPLICANT:  FUJII, Toshio
CC       APPLICANT:  IWAMATSU, Akihiro
CC       APPLICANT:  YOSHIMOTO, Hiroyuki
CC       APPLICANT:  MINEIOKI, Toshitaka
CC       APPLICANT:  BOGAKI, Takayuki
CC       APPLICANT:  NAGASAWA, Naoshi
CC       TITLE OF INVENTION:  ALCOHOL ACETYLTRANSFERASE GENES AND USE
CC       TITLE OF INVENTION:  THEREOF
CC       NUMBER OF SEQUENCES:  19
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Foley & Lardner
CC         STREET:  3000 K Street, N.W., Suite 500
CC         CITY:  Washington, D.C.
CC         COUNTRY:  USA
CC         ZIP:  20007-5109
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/465,334
CC         FILING DATE:  05-JUN-1995
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 08/077,939
CC         FILING DATE:  18-JUN-1993
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  JP 62997/1993
CC         FILING DATE:  26-FEB-1993
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  JP 184328/1992
CC         FILING DATE:  18-JUN-1992
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  BENT, Stephen A.
CC         REGISTRATION NUMBER:  29,768
CC         REFERENCE/DOCKET NUMBER:  49441/101 KYPA
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (202)672-5300
CC         TELEFAX:  (202)672-5399
CC         TELEX:  904136
CC     INFORMATION FOR SEQ ID NO:  15:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  525 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   525 AA; 61062 MW; 1464200 CN;

  Query Match           26.8%;  Score 67;  DB 1;  Length 525;
  Best Local Similarity  25.0%;  Pred. No. 1.02e+02;
  Matches     5;  Conservative    10;  Mismatches    5;  Indels    0;  Gaps    0;

Db     106 HDYISVLQELKLSGVVLNEQ 125
           : ::::|::| |: : | :|
Qy      15 RSFLAILHQLTLTIMSLVDQ 34


RESULT   14
ID   US-08-461-621-15     STANDARD;     PRT;   525 AA.
XX
AC   XXXXXX
XX
DT
XX
DE   Sequence 15, Application US/08461621
XX
CC    Sequence 15, Application US/08461621
CC    Patent No. 5686284
CC     GENERAL INFORMATION:
CC       APPLICANT:  FUJII, Toshi
CC       APPLICANT:  IWAMATSU, Akihiro
CC       APPLICANT:  YOSHIMOTO, Hiroyuki
CC       APPLICANT:  MINEIOKI, Toshitaka
CC       APPLICANT:  BOGAKI, Takayuki
CC       APPLICANT:  NAGASAWA, Naoshi
CC       TITLE OF INVENTION:  ALCOHOL ACETYLTRANSFERASE
CC       TITLE OF INVENTION:  THEREOF
CC       NUMBER OF SEQUENCES:  19
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Foley & Lardner
CC         STREET:  3000 K Street, N.W., Suite 500
CC         CITY:  Washington, D.C.
CC         COUNTRY:  USA
CC         ZIP:  20007-5109
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/461,621
CC         FILING DATE:  05-JUN-1995
CC         CLASSIFICATION:  424
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 08/077,939
CC         FILING DATE:  18-JUN-1993
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  JP 62997/1993
CC         FILING DATE:  26-FEB-1993
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  JP 184328/1992
CC         FILING DATE:  18-JUN-1992
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  BENT, Stephen A.
CC         REGISTRATION NUMBER:  29,768
CC         REFERENCE/DOCKET NUMBER:  49441/103 KYPA
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (202)672-5300
CC         TELEFAX:  (202)672-5399
CC         TELEX:  904136
CC     INFORMATION FOR SEQ ID NO:  15:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  525 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   525 AA; 61062 MW; 1464200 CN;

  Query Match           26.8%;  Score 67;  DB 1;  Length
  Best Local Similarity  25.0%;  Pred. No. 1.02e+02;
  Matches     5;  Conservative    10;  Mismatches    5;

Db     106 HDYISVLQELKLSGVVLNEQ 125
           : ::::|::| |: : | :|
Qy      15 RSFLAILHQLTLTIMSLVDQ 34


RESULT   15
ID   US-08-077-939-17     STANDARD;     PRT;   525 AA.
XX
AC   XXXXXX
XX
DT
XX
DE   Sequence 17, Application US/08077939
XX
CC    Sequence 17, Application US/08077939
CC    Patent No. 5521088
```

```
CC      GENERAL INFORMATION:
CC        APPLICANT:  FUJII, Toshio
CC        APPLICANT:  IWAMATSU, Akihiro
CC        APPLICANT:  YOSHIMOTO, Hiroyuki
CC        APPLICANT:  MINETOKI, Toshitaka
CC        APPLICANT:  BOGAKI, Takayuki
CC        APPLICANT:  NAGASAWA, Naoshi
CC        TITLE OF INVENTION:  ALCOHOL ACETYLTRANSFERASE GENES AND USE
CC        TITLE OF INVENTION:  THEREOF
CC        NUMBER OF SEQUENCES:  19
CC        CORRESPONDENCE ADDRESS:
CC          ADDRESSEE:  Foley & Lardner
CC          STREET:  3000 K Street, N.W., Suite 500
CC          CITY:  Washington, D.C.
CC          COUNTRY:  USA
CC          ZIP:  20007-5109
CC        COMPUTER READABLE FORM:
CC          MEDIUM TYPE:  Floppy disk
CC          COMPUTER:  IBM PC compatible
CC          OPERATING SYSTEM:  PC-DOS/MS-DOS
CC          SOFTWARE:  Patentin Release #1.0, Version #1.25
CC        CURRENT APPLICATION DATA:
CC          APPLICATION NUMBER:  US/08/077,939
CC          FILING DATE:  18-JUN-1932
CC          CLASSIFICATION:  435
CC        PRIOR APPLICATION DATA:
CC          APPLICATION NUMBER:  JP 184328/1992
CC          FILING DATE:  18-JUN-1992
CC        PRIOR APPLICATION DATA:
CC          APPLICATION NUMBER:  JP 62997/1993
CC          FILING DATE:  26-FEB-1993
CC        ATTORNEY/AGENT INFORMATION:
CC          NAME:  BENT, Stephen A.
CC          REGISTRATION NUMBER:  29,768
CC          REFERENCE/DOCKET NUMBER:  49441/101 KYPA
CC        TELECOMMUNICATION INFORMATION:
CC          TELEPHONE:  (202)672-5300
CC          TELEFAX:  (202)672-5399
CC          TELEX:  904136
CC      INFORMATION FOR SEQ ID NO:  17:
CC        SEQUENCE CHARACTERISTICS:
CC          LENGTH:  525 amino acids
CC          TYPE:  amino acid
CC          TOPOLOGY:  linear
CC        MOLECULE TYPE:  protein
SQ   SEQUENCE   525 AA; 61064 MW; 1459439 CN;

  Query Match           26.8%;  Score 67;  DB 1;  Length 525;
  Best Local Similarity  25.0%;  Pred. No. 1.02e+02;
  Matches    5;  Conservative   10;  Mismatches   5;  Indels   0;  Gaps   0;

Db     106 HDYISVLQELKLSGVVLNEQ 125
           : :::|:::| |: : | :|
Qy      15 RSFLAILHQLTLTIMSLVDQ 34


Search completed: Fri Aug 25 18:06:39 2000
Job time : 9 secs.
```

MPSRCH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 25 19:03:20 2000; MasPar time 2.97 Seconds
271.560 Million cell updates/sec

Tabular output not generated.

Title: >US-09-257-585-7
Description: (1-34) from US09257585.pep
Perfect Score: 250
Sequence: 1 MRPAIPTQLYFPTDRSFLAILHQLTLTTMSLVDQ 34

Scoring table: PAM 150
Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq36
1:geneseqp

Statistics: Mean 24.037; Variance 99.637; scale 0.241

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% query Match	length	DB	ID	Description	Pred. No.
1	82	32.8	289	1	R80192	UGT1C Exon 1 product.	1.22e+01
2	77	30.8	432	1	W59848	Mature protein sequenc	3.07e+01
3	77	30.8	464	1	R42929	Human antithrombin III	3.07e+01
4	77	30.8	464	1	R42935	Human antithrombin III	3.07e+01
5	77	30.8	464	1	R42941	Human antithrombin III	3.07e+01
6	77	30.8	464	1	R42927	Human antithrombin III	3.07e+01
7	77	30.8	464	1	R42934	Human antithrombin III	3.07e+01
8	77	30.8	464	1	R42913	Human antithrombin III	3.07e+01
9	77	30.8	464	1	R42933	Human antithrombin III	3.07e+01
10	77	30.8	464	1	R42914	Human antithrombin III	3.07e+01
11	77	30.8	464	1	R42910	Human antithrombin III	3.07e+01
12	77	30.8	464	1	R42930	Human antithrombin III	3.07e+01
13	77	30.8	464	1	R10387	Antithrombin III mutan	3.07e+01
14	77	30.8	464	1	R42906	Human antithrombin III	3.07e+01
15	77	30.8	464	1	R10386	Antithrombin III mutan	3.07e+01
16	77	30.8	464	1	R42932	Human antithrombin III	3.07e+01
17	77	30.8	464	1	R42937	Human antithrombin III	3.07e+01
18	77	30.8	464	1	R42931	Human antithrombin III	3.07e+01
19	77	30.8	464	1	R42896	Human antithrombin III	3.07e+01
20	77	30.8	464	1	R42915	Human antithrombin III	3.07e+01
21	77	30.8	464	1	R42922	Human antithrombin III	3.07e+01
22	77	30.8	464	1	R42902	Human antithrombin III	3.07e+01
23	77	30.8	464	1	R42939	Human antithrombin III	3.07e+01
24	77	30.8	464	1	R42926	Human antith	3.07e+01
25	77	30.8	464	1	R42940	Human antith	3.07e+01
26	77	30.8	464	1	R42919	Human antith	3.07e+01
27	77	30.8	464	1	R42899	Human antith	3.07e+01
28	77	30.8	464	1	R42912	Human antith	3.07e+01
29	77	30.8	464	1	R42904	Human antith	3.07e+01
30	77	30.8	464	1	R42907	Human antith	3.07e+01
31	77	30.8	464	1	R42909	Human antith	3.07e+01
32	77	30.8	464	1	R42928	Human antith	3.07e+01
33	77	30.8	464	1	R10381	Antithrombi	3.07e+01
34	77	30.8	464	1	R42898	Human antith	3.07e+01
35	77	30.8	464	1	R42938	Human antith	3.07e+01
36	77	30.8	464	1	R00445	Sequence of	3.07e+01
37	77	30.8	464	1	R42921	Human antith	3.07e+01
38	77	30.8	464	1	R42901	Human antith	3.07e+01
39	77	30.8	464	1	R42903	Human antith	3.07e+01
40	77	30.8	464	1	R42905	Human antith	3.07e+01
41	77	30.8	464	1	R42911	Human antith	3.07e+01
42	77	30.8	464	1	R10389	Antithrombi	3.07e+01
43	77	30.8	464	1	R42895	Human antith	3.07e+01
44	77	30.8	464	1	R42925	Human antith	3.07e+01
45	77	30.8	464	1	R42936	Human antith	3.07e+01

ALIGNMENTS

RESULT 1
ID R80192 standard; Protein; [illegible] AA.
AC R80192;
DT 27-JAN-1999 (first entry)
DE UGT1C Exon 1 product.
KW UGT1A; UGT1BP; UGT1*; UGT1B; UGT1E; UGT1F; isozyme
KW UDP-glucuronosyl transferase; CN.
OS Homo sapiens.
FH Key Location/Qualifiers
FT misc_difference 23
FT /note= "Pro encoded by CAG!"
FT misc_difference 37
FT /note= "Asp encoded by CAT!"
FT misc_difference 40
FT /note= "Gly encoded by CAG!"
FT misc_difference 256
FT /note= "Leu encoded by CYG!"
PN WO9212987 A.
PD 06-AUG-1992.
PF 10-JAN-1992; U00282.
PR 10-JAN-1991; US-638455.
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
PI Owens IS, Ritter JK;
DR WPI; 92-284592/34.
DR N-PSDB; Q33022.
PT Isolated gene locus UGT1, DNA segments and diagno
PT for diagnosing Gilbert's disease and Crigler Naji
PT types I and II
PS Disclosure; Fig 1b; 99pp; English.
CC In order to obtain this amino acid sequence, base
CC inserted between bases T978 and G979 of the cDNA
CC sequence of Q33022.
CC The isolated gene locus, UGT1, has a sequence of
CC which represent (1) Exon 1, comprising 6 transcri
CC (UGT1F, E, D, C, BP and A), represented in Q27998
CC Q30020-24 respectively;
CC (2) Exon 2, represented in Q33025;
CC (3) Exon 3, represented in Q33026;
CC (4) Exon 4, represented in Q33026;
CC (5) Exon 5, represented in Q33027; and
CC (6) about 69 kb of non-sequenced DNA.
CC Six unique N-termini of 289-289 amino acids are co
CC the six different first exons and identical C-term
CC acids are encoded by the common exons 2-5. The
CC encodes a family of UDP-glucuronosyl transferase
CC which metabolise bilirubin.
CC Patients having Crigler-Najjar Syndrome (CN) type

```
CC   mutation present in the second common exon.
SQ   Sequence   289 AA;

  Query Match           42.8%;  Score 92;  DB 1;  Length 289;
  Best Local Similarity  46.4%;  Pred. No. 1.22e+01;
  Matches   13;  Conservative    7;  Mismatches   5;  Indels    3,  Gaps    3,

Db      97 TQLYFPTIEH-PLKKIFFSMAMLNNMSLV 123
           ||||| |:: || ::: :: | ||||
Qy       7 TQLYFPTDRSFLA-ILHQLT-LTTMSLV 32


RESULT    2
ID   W59840 standard; Protein; 432 AA.
AC   W59840;
DT   20-NOV-1998  (first entry)
DE   Mature protein sequence of antithrombin III (ATIII).
KW   Protein expression; monocotyledon plant cell;
KW   glycosylated alpha 1-antitrypsin; AAT; glycosylated antithrombin III;
KW   ATIII; human serum albumin; HSA; subtilisin BPN'; treatment, emphysema,
KW   antithrombotic; blood replacement.
OS   Homo sapiens.
PN   WO9836085-A1.
PD   20-AUG-1998.
PF   13-FEB-1998; U03068.
PR   13-FEB-1997; US-038170.
PR   13-FEB-1997; US-037991.
PR   13-FEB-1997; US-038168.
PR   13-FEB-1997; US-038169.
PA   (PHYT-) APPLIED PHYTOLOGICS INC.
PI   Rodriguez RL, Sutliff TD;
DR   WPI; 98-467179/40.
DR   N-PSDB; V41727.
PT   Expressing mature, glycosylated proteins in monocotyledonous plant
PT   cells - from chimeric gene including signal peptide sequence,
PT   specifically therapeutic agents and industrial enzymes
PS   Disclosure; Pages 29-30; 53pp; English.
CC   The present sequence represents the mature protein of antithrombin III
CC   (ATIII). The protein is used to exemplify the invention. The
CC   specification describes a method for producing mature heterologous
CC   protein in monocotyledonous plant cells. The method comprises
CC   transforming the cells with a chimeric gene comprising a monocotyledon
CC   transcription regulator, inducible either during seed maturation or by
CC   adding/removing a small molecule, DNA encoding the heterologous protein,
CC   and DNA encoding a signal peptide, with the signal peptide causing
CC   secretion of the protein from the cell. Proteins expressed in this
CC   manner include mature glycosylated alpha 1-antitrypsin (AAT) with a
CC   glycosylation pattern that significantly increases its serum half life,
CC   mature glycosylated antithrombin III (ATIII), mature human serum albumin
CC   (HSA) having the native folding pattern as shown by bilirubin-binding
CC   characteristics, or mature active subtilisin BPN'. These proteins are
CC   useful therapeutically (e.g. AAT for treating emphysema, ATIII as
CC   antithrombotic and HSA as blood replacement) or as industrial enzymes
CC   (BPN' is used in detergents).
SQ   Sequence   432 AA;

  Query Match           30.8%;  Score 77;  DB 1;  Length 432;
  Best Local Similarity  24.2%;  Pred. No. 3.07e+01;
  Matches    8;  Conservative   15;  Mismatches  10;  Indels    0;  Gaps    0;

Db     393 RSLNPNRVTPKANRPFLVFIREVPLNTIIFMGR 425
           |:  | :: | ::|:|| ::::: | |: :: :
Qy       2 RPAIPTQLYFPTDRSFLAILHQLTLTTMSLVDQ 34


RESULT    3
ID   R42929 standard; Protein; 464 AA.
AC   R42929;
DT   13-MAY-1994  (first entry)
DE   Human antithrombin III mutant 3F5R.
KW   AT III; mutant; serine protease inhibitor; serpin; thrombosis;
KW   blood coagulation; anticoagulant; site directed mutagenesis;
KW   heparin binding site.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   peptide         1. .32
FT                   /label= signal_peptide
FT   protein         33. .464
FT                   /note= "mature wild-type AT III"
FT   region          43. .46
FT                   /note= "amino acids 11-14 of mature protein;
FT                   wild-type Lys11 and Asp14 are substd.
FT                   by Ile and Ser, respectively"
FT   region          416. .430
FT                   /note= "amino acids 384-398 of mature protein;
FT                   wild-type Ile390-Ala391 are substituted
FT                   by Ala-Leu"
PN   EP-568833-A.
PD   10-NOV-1993.
PF   08-APR-1993; 105829.
PR   10-APR-1992; JP-090488.
PR   22-FEB-1993; JP-031855.
PA   (EISA ) EISAI CO LTD.
```

```
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   peptide         1. .32
FT                   /label= signal_peptide
FT   protein         33. .464
FT                   /note= "mature wild type AT III"
FT   region          157. .165
FT                   /note= "amino acids 125-133 of mature protein;
FT                   wild-type Lys133 is substituted by Gln"
FT   region          416. .430
FT                   /note= "amino acids 384-398 of mature protein;
FT                   wild-type Ala391-Gly392 are substituted
FT                   by Phe-Pro"
PN   EP-568833-A.
PD   10-NOV-1993.
PF   08-APR-1993; 105829.
PR   10-APR-1992; JP-090488.
PR   22-FEB-1993; JP-031855.
PA   (EISA ) EISAI CO LTD.
PI   Kato H,  Mizui Y,  Nagaoka N,  Seto T,  Suzuki N,  Suzuki S;
PI   Yoshitake S;
DR   WPI; 93-352985/45.
PT   New human antithrombin III mutants of high antithrombin activity
PT   in absence of heparin - useful as anticoagulant for treating
PT   thrombotic disease
PS   Example 5; Page 92 93; 137pp; English.
CC   The invention covers mutant versions of wild type human AT III
CC   (R42895) which have at least 1 amino acid mutation in the 11-14,
CC   41-47, 125-133 and/or 384-398 regions of the mature protein. The
CC   mutants have high antithrombin activity, even in the absence of
CC   heparin and are useful as anticoagulants. Mutation of the heparin
CC   binding site prevents interaction of the mutant AT III with
CC   vascular endothelial cells; the mutants have a long in vivo
CC   half-life and are resistant to inactivation by neutrophil elastase.
SQ   Sequence   464 AA;

  Query Match           30.8%;  Score 77;  DB 1;  Length 464;
  Best Local Similarity  24.2%;  Pred. No. 3.07e+01;
  Matches    8;  Conservative   15;  Mismatches  10;  Indels    0;  Gaps    0;

Db     425 RSLNPNRVTPKANRPFLVFIREVPLNTIIFMGR 457
           |:  | :: | ::|:|| ::::: | |: :: :
Qy       2 RPAIPTQLYFPTDRSFLAILHQLTLTTMSLVDQ 34


RESULT    4
ID   R42935 standard; Protein; 464 AA.
AC   R42935;
DT   13-MAY-1994  (first entry)
DE   Human antithrombin III mutant 9G35R.
KW   AT III; mutant; serine protease inhibitor, serpin, thrombosis;
KW   blood coagulation; anticoagulant; site-directed mutagenesis;
KW   heparin binding site.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   peptide         1. .32
FT                   /label= signal_peptide
FT   protein         33. .464
FT                   /note= "mature wild-type AT III"
FT   region          43. .46
FT                   /note= "amino acids 11-14 of mature protein;
FT                   wild-type Lys11 and Asp14 are substd.
FT                   by Ile and Ser, respectively"
FT   region          416. .430
FT                   /note= "amino acids 384-398 of mature protein;
FT                   wild-type Ile390-Ala391 are substituted
FT                   by Ala-Leu"
PN   EP-568833-A.
PD   10-NOV-1993.
PF   08-APR-1993; 105829.
PR   10-APR-1992; JP-090488.
PR   22-FEB-1993; JP-031855.
PA   (EISA ) EISAI CO LTD.
```

PI Kato H, Mizui Y, Nagaoka N, Seto T, Suzuki N, Suzuki S,
PI Yoshitake S;
DR WPI; 93-352985/45.
PI New human antithrombin III mutants of high antithrombin activity
PI in absence of heparin - useful as anticoagulant for treating
PT thrombotic disease
PS Example 5; Page 101-103; 137pp; English.
CC The invention covers mutant versions of wild type human AT III
CC (R42895) which have at least 1 amino acid mutation in the 11-14,
CC 41-47, 125-133 and/or 384-398 regions of the mature protein. The
CC mutants have high antithrombin activity, even in the absence of
CC heparin and are useful as anticoagulants. Mutation of the heparin
CC binding site prevents interaction of the mutant AT III with
CC vascular endothelial cells; the mutants have a long in vivo
CC half life and are resistant to inactivation by neutrophil elastase.
SQ Sequence 464 AA;

Query Match 30.8%; Score 77; DB 1; Length 464;
Best Local Similarity 24.2%; Pred. No. 3.07e-01;
Matches 8; Conservative 15; Mismatches 10; Indels 0; Gaps 0;

Db 425 RSLNPNRVTPKANRPFLVFIREVPLNTIIFMGR 457
|: | :: | :|:|| ||||| | |: :: :
Qy 2 RPAIPTQLYFPTDRSFLAILHQLTLTTMSLVDQ 34

RESULT 5
ID R42941 standard; Protein; 464 AA.
AC R42941;
DT 13-MAY-1994 (first entry)
DE Human antithrombin III mutant 127G35R.
KW AT III; mutant; serine protease inhibitor; serpin; thrombosis;
KW blood coagulation; anticoagulant; site-directed mutagenesis;
KW heparin binding site.
OS Homo sapiens.
FH Key Location/Qualifiers
FT peptide 1. .32
FT /label= signal_peptide
FT protein 33. .464
FT /note= "mature wild-type AT III"
FT region 157. .165
FT /note= "amino acids 125-133 of mature protein;
FT wild-type Lys125, Arg129 and Arg132-Lys133
FT are substituted by Gln, Gln and Gln-Asn,
FT respectively"
FT region 416. .430
FT /note= "amino acids 384-398 of mature protein;
FT wild-type Ile390-Ala391 are substituted
FT by Ala-Leu"
PN EP 568833-A.
PD 10-NOV-1993.
PF 08-APR-1993; 105829.
PR 10-APR-1992; JP-090488.
PR 22-FEB-1993; JP-041855.
PA (EISA) EISAI CO LTD.
PI Kato H, Mizui Y, Nagaoka N, Seto T, Suzuki N, Suzuki S;
PI Yoshitake S;
DR WPI; 93-352985/45.
PT New human antithrombin III mutants of high antithrombin activity
PT in absence of heparin - useful as anticoagulant for treating
PT thrombotic disease
PS Example 5; Page 111-112; 137pp; English.
CC The invention covers mutant versions of wild-type human AT III
CC (R42895) which have at least 1 amino acid mutation in the 11-14,
CC 41-47, 125-133 and/or 384-398 regions of the mature protein. The
CC mutants have high antithrombin activity, even in the absence of
CC heparin and are useful as anticoagulants. Mutation of the heparin
CC binding site prevents interaction of the mutant AT III with
CC vascular endothelial cells; the mutants have a long in vivo
CC half-life and are resistant to inactivation by neutrophil elastase.
SQ Sequence 464 AA;

Query Match 30.8%; Score 77; DB 1; Length 464.
Best Local Similarity 24.2%; Pred. No. 3.07e-01;
Matches 8; Conservative 15; Mismatches 10; Indels 0; Gaps 0;

Db 425 RSLNPNRVTPKANRPFLVFIREVPLNTIIFMGR 457
|: | :: | :|:|| ||||| | |: :: :
Qy 2 RPAIPTQLYFPTDRSFLAILHQLTLTTMSLVDQ 34

RESULT 6
ID R42927 standard; Protein; 464 AA.
AC R42927;
DT 13-MAY-1994 (first entry)
DE Human antithrombin III mutant 127G35R.
KW AT III; mutant; serine protease inhibitor; serpin;
KW blood coagulation; anticoagulant; site-directed mu
KW heparin binding site.
OS Homo sapiens.
FH Key Location/Qualifiers
FT peptide 1. .32
FT /label= signal_peptide
FT protein 33. .464
FT /note= "mature wild-type AT III"
FT region 157. .165
FT /note= "amino acids 125-133 of mat
FT wild-type Arg132 is substituted by
FT region 416. .430
FT /note= "amino acids 384-398 of mat
FT wild-type Asp391-Gly392 is substit
FT by Phe-Ile"
PN EP-568833-A.
PD 10-NOV-1993.
PF 08-APR-1993; 105829.
PR 10-APR-1992; JP-090488.
PR 22-FEB-1993; JP-041855.
PA (EISA) EISAI CO LTD.
PI Kato H, Mizui Y, Nagaoka N, Seto T, Suzuki N,
PI Yoshitake S;
DR WPI; 93-352985/45.
PT New human antithrombin III mutant's of high antith
PT in absence of heparin - useful as anticoagulant fo
PT thrombotic disease
PS Example 5, Page 89-90; 137pp; English.
CC The invention covers mutant versions of wild-type
CC (R42895) which have at least 1 amino acid mutation
CC 41-47, 125-133 and/or 384-398 regions of the mature
CC mutants have high antithrombin activity, even in t
CC heparin and are useful as anticoagulants. Mutation
CC binding site prevents interaction of the mutant AT
CC vascular endothelial cells; the mutants have a lon
CC half-life and are resistant to inactivation by neu
SQ Sequence 464 AA;

Query Match 30.8%; Score 77; DB 1; Length
Best Local Similarity 24.2%; Pred. No. 3.07e-01;
Matches 8; Conservative 15; Mismatches 10; Indels

Db 425 RSLNPNRVTPKANRPFLVFIREVPLNTIIFMGR 457
|: | :: | :|:|| ||||| | |: :: :
Qy 2 RPAIPTQLYFPTDRSFLAILHQLTLTTMSLVDQ 34

RESULT 7
ID R42934 standard; Protein; 464 AA.
AC R42934;
DT 13-MAY-1994 (first entry)
DE Human antithrombin III mutant 96G5R.
KW AT III; mutant; serine protease inhibitor; serpin;
KW blood coagulation; anticoagulant; site-directed mu
KW heparin binding site.
OS Homo sapiens.
FH Key Location/Qualifiers
FT peptide 1. .32
FT /label= signal_peptide

```
FT   protein          33. .464
FT                    /note= "mature wild-type AT III"
FT   region           43. .46
FT                    /note= "amino acids 11 14 of mature protein,
FT                    wild-type Lys11 and Asp14 are substd.
FT                    by Ile and Ser, respectively"
FT   region           416. .430
FT                    /note= "amino acids 384-398 of mature protein;
FT                    wild-type Ala391-Gly392 is substituted
FT                    by Val-Pro"
PN   EP-568833-A.
PD   10-NOV-1993.
PF   08-APR-1993; 105829.
PR   10-APR-1992; JP-090488.
PR   22-FEB-1993; JP-031855.
PA   (EISA ) EISAI CO LTD.
PI   Kato H,  Mizui Y,  Nagaoka N,  Seto T,  Suzuki N,  Suzuki S;
PI   Yoshitake S;
DR   WPI; 93-352985/45.
PT   New human antithrombin III mutants of high antithrombin activity
PT   in absence of heparin - useful as anticoagulant for treating
PT   thrombotic disease
PS   Example 5; Page 100-101; 137pp; English.
CC   The invention covers mutant versions of wild type human AT III
CC   (R42895) which have at least 1 amino acid mutation in the 11-14,
CC   41-47, 125-133 and/or 384-398 regions of the mature protein. The
CC   mutants have high antithrombin activity, even in the absence of
CC   heparin and are useful as anticoagulants. Mutation of the heparin
CC   binding site prevents interaction of the mutant AT III with
CC   vascular endothelial cells; the mutants have a long in vivo
CC   half-life and are resistant to inactivation by neutrophil elastase.
SQ   Sequence   464 AA;

  Query Match             30.8%;  Score 77;  DB 1;  Length 464;
  Best Local Similarity   24.2%;  Pred. No. 3.07e+01;
  Matches    8;  Conservative   15;  Mismatches  10;  Indels   0;  Gaps   0;

Db    425 RSLNPNRVTPKANRPFLVFIREVPLNTIIFMGR 457
          |:  | :: | ::|:|| ::::: | |: :: :
Qy      2 RPAIPTQLYFPTDRSFIAIHQITLTTMSLVDQ 34


RESULT    8
ID   R42913 standard; Protein; 464 AA.
AC   R42913;
DT   13-MAY-1994  (first entry)
DE   Human antithrombin III mutant 9R.
KW   AT III; mutant; serine protease inhibitor; serpin; thrombosis;
KW   blood coagulation; anticoagulant; site directed mutagenesis;
KW   heparin binding site.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   peptide          1. .32
FT                    /label= signal_peptide
FT   protein          33. .464
FT                    /note= "mature wild-type AT III"
FT   region           416. .430
FT                    /note= "amino acids 384-398 of mature protein;
FT                    wild-type Ala384 is substituted by Gly"
PN   EP-568833-A.
PD   10-NOV-1993.
PF   08-APR-1993; 105829.
PR   10-APR-1992; JP-090488.
PR   22-FEB-1993; JP-031855.
PA   (EISA ) EISAI CO LTD.
PI   Kato H,  Mizui Y,  Nagaoka N,  Seto T,  Suzuki N,  Suzuki S;
PI   Yoshitake S;
DR   WPI; 93-352985/45.
PT   New human antithrombin III mutants of high antithrombin activity
PT   in absence of heparin - useful as anticoagulant for treating
PT   thrombotic disease
PS   Example 3; Page 66-68; 137pp; English.
CC   The invention covers mutant versions of wild-type human AT III
CC   (R42895) which have at least 1 amino acid mutation in the 11-14,
CC   41-47, 125-133 and/or 384-398 regions of the mature protein. The
CC   mutants have high antithrombin activity, even in the absence of
CC   heparin and are useful as anticoagulants. Mutation of the heparin
CC   binding site prevents interaction of the mutant AT III with
CC   vascular endothelial cells; the mutants have a long in vivo
CC   half-life and are resistant to inactivation by neutrophil elastase.
SQ   Sequence   464 AA;

  Query Match             30.8%;  Score 77;  DB 1;  Length 464;
  Best Local Similarity   24.2%;  Pred. No. 3.07e+01;
  Matches    8;  Conservative   15;  Mismatches  10;  Indels   0;  Gaps   0;

Db    425 RSLNPNRVTPKANRPFLVFIREVPLNTIIFMGR 457
          |:  | :: | ::|:|| ::::: | |: :: :
Qy      2 RPAIPTQLYFPTDRSFLAILHQLTLTTMSLVDQ 34


RESULT    9
ID   R42933 standard; Protein; 464 AA.
AC   R42933;
DT   13-MAY-1994  (first entry)
DE   Human antithrombin III mutant 9G5R.
KW   AT III; mutant, serine protease inhibitor, serpin; thrombosis;
KW   blood coagulation, anticoagulant, site directed mutagenesis;
KW   heparin binding site.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   peptide          1. .32
FT                    /label= signal_peptide
FT   protein          33. .464
FT                    /note= "mature wild-type AT III"
FT   region           43. .46
FT                    /note= "amino acids 11-14 of mature protein;
FT                    wild-type Lys11 and Asp14 are substd.
FT                    by Ile and Ser, respectively"
FT   region           416. .430
FT                    /note= "amino acids 384-398 of mature protein;
FT                    wild-type Ala391-Gly392 is substituted
FT                    by Phe-Pro"
PN   EP-568833-A.
PD   10-NOV-1993.
PF   08-APR-1993; 105829.
PR   10-APR-1992; JP-090488.
PR   22-FEB-1993; JP-031855.
PA   (EISA ) EISAI CO LTD.
PI   Kato H,  Mizui Y,  Nagaoka N,  Seto T,  Suzuki N,  Suzuki S;
PI   Yoshitake S;
DR   WPI; 93-352985/45.
PT   New human antithrombin III mutants of high antithrombin activity
PT   in absence of heparin - useful as anticoagulant for treating
PT   thrombotic disease
PS   Example 5; Page 98-99; 137pp; English.
CC   The invention covers mutant versions of wild-type human AT III
CC   (R42895) which have at least 1 amino acid mutation in the 11-14,
CC   41-47, 125-133 and/or 384-398 regions of the mature protein. The
CC   mutants have high antithrombin activity, even in the absence of
CC   heparin and are useful as anticoagulants. Mutation of the heparin
CC   binding site prevents interaction of the mutant AT III with
CC   vascular endothelial cells; the mutants have a long in vivo
CC   half-life and are resistant to inactivation by neutrophil elastase.
SQ   Sequence   464 AA;

  Query Match             30.8%;  Score 77;  DB 1;  Length 464;
  Best Local Similarity   24.2%;  Pred. No. 3.07e+01;
  Matches    8;  Conservative   15;  Mismatches  10;  Indels   0;  Gaps   0;

Db    425 RSLNPNRVTPKANRPFLVFIREVPLNTIIFMGR 457
          |:  | :: | ::|:|| ::::: | |: :: :
Qy      2 RPAIPTQLYFPTDRSFLAILHQLTLTTMSLVDQ 34


RESULT   10
```

```
ID   R42914 standard; Protein; 464 AA.
AC   R42914;
DT   13-MAY-1994  (first entry)
DE   Human antithrombin III mutant 19R.
KW   AT III; mutant; serine protease inhibitor; serpin; thrombosis;
KW   blood coagulation; anticoagulant; site-directed mutagenesis;
KW   heparin binding site.
OS   Homo sapiens.
FH   Key             Location/qualifiers
FT   peptide         1. .32
FT                   /label= signal_peptide
FT   protein         33. .464
FT                   /note= "mature wild-type AT III"
FT   region          416. .430
FT                   /note= "amino acids 384-398 of mature protein;
FT                   Wild-type Val389 is substituted by Pro"
PN   EP-568833-A.
PD   10-NOV-1993.
PF   08-APR-1993; 105829.
PR   10-APR-1992; JP-090488.
PR   22-FEB-1993; JP-031855.
PA   (EISA ) EISAI CO LTD.
PI   Kato H  Mizui Y  Nagaoka N  Seto T  Suzuki N  Suzuki S;
PI   Yoshitake S;
DR   WPI; 93-352985/45.
PT   New human antithrombin III mutants of high antithrombin activity
PT   in absence of heparin - useful as anticoagulant for treating
PT   thrombotic disease
PS   Example 4; Page 68-69; 137pp; English.
CC   The invention covers mutant versions of wild-type human AT III
CC   (R42895) which have at least 1 amino acid mutation in the 11-14,
CC   41-47, 125-133 and/or 384-398 regions of the mature protein. The
CC   mutants have high antithrombin activity, even in the absence of
CC   heparin and are useful as anticoagulants. Mutation of the heparin
CC   binding site prevents interaction of the mutant AT III with
CC   vascular endothelial cells; the mutants have a long in vivo
CC   half-life and are resistant to inactivation by neutrophil elastase.
SQ   Sequence   464 AA;

  Query Match             30.8%;  Score 77;  DB 1;  Length 464
  Best Local Similarity  24.2%;  Pred. No. 3.07e-01;
  Matches     8;  Conservative   15;  Mismatches  10;  Indels   0;  Gaps   0;

Db     425 RSLNPNRVTFKANRPFLVFIREVPLNTIIFMGR 457
           |:  | :: \ ::|:|| ::::: | |: :: :
Qy       2 RPAIPIQLYFPIDRSFLAILHQLILITMSLVDQ 34


RESULT   11
ID   R42910 standard; Protein; 464 AA.
AC   R42910;
DT   13-MAY-1994  (first entry)
DE   Human antithrombin III mutant 34R.
KW   AT III; mutant; serine protease inhibitor; serpin; thrombosis;
KW   blood coagulation; anticoagulant; site-directed mutagenesis;
KW   heparin binding site.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   peptide         1. .32
FT                   /label= signal_peptide
FT   protein         33. .464
FT                   /note= "mature wild-type AT III"
FT   region          416. .430
FT                   /note= "amino acids 384-398 of mature protein;
FT                   wild-type Ile390-Ala391 are substituted
FT                   by Ala-Ile"
PN   EP-568833-A.
PD   10-NOV-1993.
PF   08-APR-1993; 105829.
PR   10-APR-1992; JP-090488.
PR   22-FEB-1993; JP-031855.
PA   (EISA ) EISAI CO LTD.
PI   Kato H,  Mizui Y,  Nagaoka N,  Seto T,  Suzuki N,  Suzuki S;
PI   Yoshitake S;
DR   WPI; 93-352985/45.
PT   New human antithrombin III mutants of high antithrombin activity
PT   in absence of heparin - useful as anticoagulant for treating
PT   thrombotic disease
PS   Example 3; Page 61-63; 137pp; English.
CC   The invention covers mutant versions of wild-type human AT III
CC   (R42895) which have at least 1 amino acid mutation in the 11-14,
CC   41-47, 125-133 and/or 384-398 regions of the mature protein. The
CC   mutants have high antithrombin activity, even in the absence of
CC   heparin and are useful as anticoagulants. Mutation of the heparin
CC   binding site prevents interaction of the mutant AT III with
CC   vascular endothelial cells; the mutants have a long in vivo
CC   half-life and are resistant to inactivation by neutrophil elastase.
SQ   Sequence   464 AA;

  Query Match             30.8%;  Score 77;  DB 1;  Length 464
  Best Local Similarity  24.2%;  Pred. No. 3.07e-01;
  Matches     8;  Conservative   15;  Mismatches  10;  Indels   0;  Gaps   0;

Db     425 RSLNPNRVTFKANRPFLVFIREVPLNTIIFMGR 457
           |:  | :: \ ::|:|| ::::: | |: :: :
Qy       2 RPAIPIQLYFPIDRSFLAILHQLILITMSLVDQ 34


RESULT   12
ID   R42930 standard; Protein; 464 AA.
AC   R42930;
DT   13-MAY-1994  (first entry)
DE   Human antithrombin III mutant 76R.
KW   AT III; mutant, serine protease inhibitor; serpin;
KW   blood coagulation; anticoagulant; site-directed mu
KW   heparin binding site.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   peptide         1. .32
FT                   /label= signal_peptide
FT   protein         33. .464
FT                   /note= "mature wild-type AT III"
FT   region          157. .[illegible]
FT                   /note= "amino acids 125-133 of mat
FT                   wild-type Arg129 is substituted by
FT   region          416. .430
FT                   /note= "amino acids 384-398 of mat
FT                   wild-type Ala391-Gly392 are substi
FT                   by Phe-Pro"
PN   EP-568833-A.
PD   10-NOV-1993.
PF   08-APR-1993; 105829.
PR   10-APR-1992; JP-090488.
PR   22-FEB-1993; JP-031855.
PA   (EISA ) EISAI CO LTD.
PI   Kato H,  Mizui Y,  Nagaoka N,  Seto T,  Suzuki N,
PI   Yoshitake S;
DR   WPI; 93-352985/45.
PT   New human antithrombin III mutants of high antithro
PT   in absence of heparin - useful as anticoagulant for
PT   thrombotic disease
PS   Example 5; Page 93-95; 137pp; English.
CC   The invention covers mutant versions of wild type h
CC   (R42895) which have at least 1 amino acid mutation
CC   41-47, 125-133 and/or 384-398 regions of the mature
CC   mutants have high antithrombin activity, even in th
CC   heparin and are useful as anticoagulants. Mutation
CC   binding site prevents interaction of the mutant AT
CC   vascular endothelial cells; the mutants have a long
CC   half-life and are resistant to inactivation by neut
SQ   Sequence   464 AA;

  Query Match             30.8%;  Score 77;  DB 1;  Length
  Best Local Similarity  24.2%;  Pred. No. 3.07e-01;
  Matches     8;  Conservative   15;  Mismatches  10;  I
```

```
Db     425 RSLNPNRVTPKANRPPLVFIREVPLNTIIFMGR 457
           |:  | :: | ::|:|| ::::: | |: :: :
Qy       2 RPAIPTQLYFPTDRSFLAILHQLTLTTMSLVDQ 34


RESULT   13
ID   R10387 standard; Protein; 464 AA.
AC   R10387;
DT   10-APR-1991  (first entry)
DE   Antithrombin III mutant #8.
KW   antithrombin 3; P-region variants; heparin-dependent; Factor IIa;
KW   Factor Xa; melzothrombin; blood coagulation.
FH   Key             Location/Qualifiers
FT   region          415. .436
FT                   /label= P-region
FT                   /note= "substitutions occur in this region which is
FT                   positions 383-404 in mature ATIII"
PN   WO9100291-A.
PD   10-JAN-1991.
PF   23-JUN-1990; E01026.
PR   26-JUN-1989; EP-201675.
PA   (ALKU ) AKZO NV.
PI   Dijkema R, Visser A;
DR   WPI; 91-036710/05.
PT   New modified antithrombin III variants - with altered
PT   heparin-dependent effect to factors IIa and Xa
PS   Claim 2; Fig 2; 24pp; English.
CC   Mutations introduced into the ATIII reactive centre P-region are
CC   derived from the human prothrombin gene. Sequences known to be
CC   physiological substrates for Factor Xa were selected and used to
CC   replace the wild-type PvuII-StuI fragment of the original ATIII
CC   cDNA sequence (see Q10323). The only substitution is Ala(391) to
CC   Asp. (Numbers refer to mature ATIII). The substitution results in
CC   a shift in the heparin-dependent inhibition profile of ATIII
CC   towards factors IIa and Xa.
CC   See also R10380-6 and R10388-R10390.
SQ   Sequence   464 AA;

  Query Match           30.8%;  Score 77;  DB 1;  Length 464;
  Best Local Similarity  24.2%;  Pred. No. 3.07e+01;
  Matches     8;  Conservative   15;  Mismatches  10;  Indels   0;  Gaps   0;

Db     425 RSLNPNRVTPKANRPPLVFIREVPLNTIIFMGR 457
           |:  | :: | ::|:|| ::::: | |: :: :
Qy       2 RPAIPTQLYFPTDRSFLAILHQLTLTTMSLVDQ 34


RESULT   14
ID   R42906 standard; Protein; 464 AA.
AC   R42906;
DT   13-MAY-1994  (first entry)
DE   Human antithrombin III mutant 48R.
KW   AT III; mutant; serine protease inhibitor; serpin; thrombosis;
KW   blood coagulation; anticoagulant; site-directed mutagenesis;
KW   heparin binding site.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   peptide         1. .32
FT                   /label= signal_peptide
FT   protein         33. .464
FT                   /note= "mature wild-type AT III"
FT   region          416. .430
FT                   /note= "amino acids 384-398 of mature protein;
FT                   wild-type Ile390-Ala391-Gly392 are substd.
FT                   by Ala-Tyr-Pro"
PN   EP-568833-A.
PD   10-NOV-1993.
PF   08-APR-1993; 105829.
PR   10-APR-1992; JP-090488.
PR   22-FEB-1993; JP-031855.
PA   (EISA ) EISAI CO LTD.
PI   Kato H, Mizui Y, Nagaoka N, Sato T, Suzuki M, Suzuki S;
PI   Yoshitake S;
DR   WPI; 93-352985/45.
PT   New human antithrombin III mutants of high antithrombin activity
PT   in absence of heparin - useful as anticoagulant for treating
PT   thrombotic disease
PS   Example 3; Page 55-56; 137pp; English.
CC   The invention covers mutant versions of wild type human AT III
CC   (R42895) which have at least 1 amino acid mutation in the 11-14,
CC   41-47, 125-133 and/or 384-398 regions of the mature protein. The
CC   mutants have high antithrombin activity, even in the absence of
CC   heparin and are useful as anticoagulants. Mutation of the heparin
CC   binding site prevents interaction of the mutant AT III with
CC   vascular endothelial cells; the mutants have a long in vivo
CC   half-life and are resistant to inactivation by neutrophil elastase.
SQ   Sequence   464 AA;

  Query Match           30.8%;  Score 77;  DB 1;  Length 464;
  Best Local Similarity  24.2%;  Pred. No. 3.07e+01;
  Matches     8;  Conservative   15;  Mismatches  10;  Indels   0;  Gaps   0;

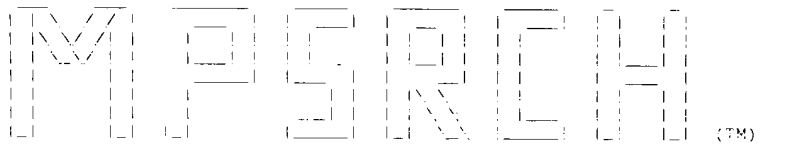
Db     425 RSLNPNRVTPKANRPPLVFIREVPLNTIIFMGR 457
           |:  | :: | ::|:|| ::::: | |: :: :
Qy       2 RPAIPTQLYFPTDRSFLAILHQLTLTTMSLVDQ 34


RESULT   15
ID   R10386 standard; Protein; 464 AA.
AC   R10386;
DT   10-APR-1991  (first entry)
DE   Antithrombin III mutant #7.
KW   antithrombin 3; P-region variants; heparin-dependent; Factor IIa;
KW   Factor Xa; melzothrombin; blood coagulation.
FH   Key             Location/Qualifiers
FT   region          415. .436
FT                   /label= P-region
FT                   /note= "substitutions occur in this region which is
FT                   positions 383-404 in mature ATIII"
PN   WO9100291-A.
PD   10-JAN-1991.
PF   23-JUN-1990; E01026.
PR   26-JUN-1989; EP 201675.
PA   (AKZO ) AKZO NV.
PI   Dijkema R, Visser A;
DR   WPI; 91-036710/05.
PT   New modified antithrombin III variants - with altered
PT   heparin-dependent effect to factors IIa and Xa
PS   Claim 2; Fig 2; 24pp; English.
CC   Mutations introduced into the ATIII reactive centre P-region are
CC   derived from the human prothrombin gene. Sequences known to be
CC   physiological substrates for Factor Xa were selected and used to
CC   replace the wild-type PvuII-StuI fragment of the original ATIII
CC   cDNA sequence (see Q10323). The substitutions are Ala(391) to Asp
CC   and Asn(396) to Glu. (Numbers refer to mature ATIII). The
CC   substitutions result in a shift in the heparin dependent inhibition
CC   profile of ATIII towards factors IIa and Xa.
CC   See also R10380-5 and R10387-R10390.
SQ   Sequence   464 AA;

  Query Match           30.8%;  Score 77;  DB 1;  Length 464;
  Best Local Similarity  24.2%;  Pred. No. 3.07e+01;
  Matches     8;  Conservative   15;  Mismatches  10;  Indels   0;  Gaps   0;

Db     425 RSLEPNRVTPKANRPPLVFIREVPLNTIIFMGR 457
           |:  | :: | ::|:|| ::::: | |: :: :
Qy       2 RPAIPTQLYFPTDRSFLAILHQLTLTTMSLVDQ 34


Search completed: Fri Aug 25 18:03:38 2000
Job time : 18 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 25 18:04:45 2000; MusPar time 3.09 Seconds
341.002 Million cell updates/sec

Tubular output not generated.

Title: >US-09-257-585-7
Description: (1-34) from US09257585.pep
Perfect Score: 250
Sequence: 1 MRPAIPTQLYFPTDRSFLAILHQLTITTMSLVDQ 34

Scoring table: PAM 150
Gap 11

Searched: 85661 seqs, 30989116 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 45.072; Variance 59.488; scale 0.590

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	82	32.8	534	1	UD13_HUMAN	UDP-GLUCURONOSYLTRANSF	2.16e-01
2	81	32.4	452	1	URIC_DROME	URICASE (EC 1.7.3.3) (	3.08e-01
3	80	32.0	327	1	DHB1_HUMAN	ESTRADIOL 17 BETA DEHY	4.39e-01
4	80	32.0	465	1	AN13_MOUSE	ANTITHROMBIN-III PRECU	4.39e-01
5	77	30.8	418	1	NUIM_DASNO	NADH-UBIQUINONE OXIDOR	1.24e+00
6	77	30.8	366	1	BCPA_DROAF	BACTERIOCHLOROPHYLL A	1.24e+00
7	77	30.8	464	1	AN13_HUMAN	ANTITHROMBIN-III PRECU	1.24e+00
8	77	30.8	887	1	MCM2_DROME	DNA REPLICATION LICENS	1.24e+00
9	76	30.4	442	1	URIC_DROVI	URICASE (EC 1.7.3.3) (	1.75e+00
10	75	30.0	379	1	CYB_SORCA	CYTOCHROME B.	2.45e+00
11	75	30.0	379	1	CYB_SPERI	CYTOCHROME B.	2.45e+00
12	75	30.0	379	1	CYB_MOSMO	CYTOCHROME B.	2.45e+00
13	75	30.0	532	1	EAT2_CAEEL	PUTATIVE SODIUM-DEPEND	2.45e+00
14	74	29.6	242	1	RPIA_CHLTR	RIBOSE 5-PHOSPHATE ISO	3.43e+00
15	74	29.6	379	1	CYB_CRAGG	CYTOCHROME B.	3.43e+00
16	74	29.6	379	1	CYB_CRAGR	CYTOCHROME B.	3.43e+00
17	74	29.6	522	1	GTK2_RAT	GLUCOSE TRANSPORTER TY	3.43e+00
18	74	29.6	567	1	ACH1_DROME	ACETYLCHOLINE RECEPTOR	3.43e+00
19	74	29.6	690	1	42_HUMAN	ERYTHROCYTE MEMBRANE P	3.43e+00
20	73	29.2	251	1	Y393_CHLTR	HYPOTHETICAL PROTEIN C	4.78e+00
21	73	29.2	365	1	BCPA_CHLTE	BACTERIOCHLOROPHYLL A	4.78e+00
22	73	29.2	379	1	CYB_MOSLE	CYTOCHROME B.	4.78e+00
23	73	29.2	379	1	CYB_DICBI	CYTOCHROME B.	4.78e+00
24	73	29.2	465	1	AN13_SHEEP	ANTITHROMBIN-III PRECU	4.78e+00
25	73	29.2	528	1	GTK2_MOUSE	GLUCOSE TRANSPORTER TY	4.78e+00
26	72	28.8	344	1	DHB1_RAT	ESTRADIOL 17 BETA DEHY	6.24e+00
27	72	28.8	344	1	DHB1_MOUSE	ESTRADIOL 17 BETA DEHY	6.24e+00
28	72	28.8	379	1	CYB_CRAGY	CYTOCHROME B.	6.24e+00
29	72	28.8	379	1	CYB_CRAGE	CYTOCHROME B.	6.24e+00
30	72	28.8	429	1	RNE_SPITH	RIBONUCLEASE	6.24e+00
31	72	28.8	434	1	ANT3_BOVIN	ANTITHROMBIN-III PRECU	6.24e+00
32	72	28.8	476	1	SYGB_BACSU	GLYCYL-TRNA	6.24e+00
33	71	28.4	333	1	ILVC_MYCAV	KETOL-ACID	8.15e+00
34	71	28.4	343	1	ILVC_MYCLE	KETOL-ACID	8.15e+00
35	71	28.4	364	1	BCPA_CHLLT	BACTERIOCHL	8.15e+00
36	71	28.4	379	1	CYB_LEMCA	CYTOCHROME B.	8.15e+00
37	71	28.4	379	1	CYB_LONFE	CYTOCHROME B.	8.15e+00
38	71	28.4	436	1	V55B_HSV7J	PROTEIN US5	8.15e+00
39	71	28.4	601	1	NUIM_DASNO	NADH-UBIQUIN	8.15e+00
40	71	28.4	766	1	COMP_BACSU	SENSOR PROT	8.15e+00
41	70	28.0	421	1	NUIM_ALLMI	NADH-UBIQUIN	1.06e+01
42	70	28.0	379	1	CYB_CRAME	CYTOCHROME B.	1.06e+01
43	70	28.0	479	1	CYB_CERRI	CYTOCHROME B.	1.06e+01
44	70	28.0	841	1	YEBL_ECOLI	PUTATIVE PE	1.06e+01
45	70	28.0	1073	1	RAS1_ONCMY	V(D)J RECOM	1.06e+01

ALIGNMENTS

RESULT 1
ID UD13_HUMAN STANDARD; PRT; 534 AA.
AC P35503;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE UDP-GLUCURONOSYLTRANSFERASE 1-3 PRECURSOR, MICROSO
DE (UDPGT) (UGT-1C) (UGT1*3) (UGT1.03) (UGT1.3) (UGT1
GN UGT1 OR GNT1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrat
OC Mammalia; Eutheria; Primates; Catarrhini; Hominid
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92147680.
RA Ritter J.K., Chen F., Sheen Y.T., Tran H.M., Kimur
RA Owens I.S.;
RT "A novel complex locus UGT1 encodes human bili
RT other UDP-glucuronosyltransferase isozymes wi
RT termini.";
RL J. Biol. Chem. 267:
CC -!- FUNCTION: UDPGT IS OF MAJOR IMPORTANCE IN THE
CC SUBSEQUENT ELIMINATION OF POTENTIALLY TOXIC
CC ENDOGENOUS COMPOUNDS.
CC -!- CATALYTIC ACTIVITY: UDP-GLUCURONATE + ACCEPTO
CC BETA-D-GLUCORONOSIDE.
CC -!- SUBCELLULAR LOCATION: MICROSOMAL.
CC -!- ALTERNATIVE PRODUCTS: THE UGT1 GENE ENCODES F
CC UDPGT ISOZYMES WHICH HAVE A DIFFERENT N-TERMIN
CC COMMON C-TERMINAL DOMAIN OF 245 RESIDUES.
CC -!- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFE
CC ----------
CC This SWISS-PROT entry is copyright. It is provided
CC between the Swiss Institute of Bioinformatics an
CC the European Bioinformatics Institute. There are
CC use by non-profit institutions as long as its
CC modified and this statement is not removed. Usage
CC entities requires a license agreement (See http://
CC or send an email to license@isb-sib.ch).
CC ----------
DR EMBL; M84127; AAA92023.1;
DR EMBL; M84124; AAA61247.1; ALT_SEQ.
DR EMBL; M84122; AAA61247.1; JOINED.
DR EMBL; M84123; AAA61247.1; JOINED.
DR MIM; 191740; -.
DR PFAM; PF00201; UDPGT; 1.
DR PROSITE; PS00375; UDPGT; 1.

```
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal;
KW   Multigene family; Microsome; Alternative splicing.
FT   SIGNAL        1     28       POTENTIAL.
FT   CHAIN        29    534       UDP-GLUCURONOSYLTRANSFERASE 1-3.
FT   TRANSMEM    492    508       POTENTIAL.
FT   CARBOHYD     119    119       POTENTIAL.
FT   CARBOHYD     142    142       POTENTIAL.
FT   CARBOHYD     296    296       POTENTIAL.
FT   CARBOHYD     348    348       POTENTIAL.
SQ   SEQUENCE   534 AA;  60338 MW;  9C5833652A4D9B3D CRC64;

  Query Match           32.8%;  Score 82;  DB 1;  Length 534;
  Best Local Similarity  46.4%;  Pred. No. 2.16e-01;
  Matches   13;  Conservative    7;  Mismatches   5;  Indels   3;  Gaps   3;

Db      97 TQLYFPTDRSFLA-ILHQLT-LTTMSLV 123
           ||||| |:: || ::: :: | ||||
Qy       7 TQLYFPTDRSFLA-ILHQLT-LTTMSLV 32


RESULT    2
ID   URIC_DROME     STANDARD;      PRT;   352 AA.
AC   P16163;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   URICASE (EC 1.7.3.3) (URATE OXIDASE).
GN   URO OR UO.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CANTON-S; TISSUE=SALIVARY GLAND;
RX   MEDLINE; 90377201.
RA   Wallrath L.L., Burnett J.B., Friedman T.B.;
RT   "Molecular characterization of the Drosophila melanogaster urate
RT   oxidase gene, an ecdysone-repressible gene expressed only in the
RT   malpighian tubules.";
RL   Mol. Cell. Biol. 10:5114-5127(1990).
CC   -!- CATALYTIC ACTIVITY: URATE + O(2) = UNIDENTIFIED PRODUCTS (THE
CC       INITIAL PRODUCTS DECOMPOSE TO FORM ALLANTOIN).
CC   -!- ENZYME REGULATION: REPRESSED BY 20-HYDROXYECDYSONE.
CC   -!- SUBCELLULAR LOCATION: PEROXISOMAL.
CC   -!- TISSUE SPECIFICITY: MALPIGHIAN TUBULES.
CC   -!- DEVELOPMENTAL STAGE: THIRD INSTAR LARVAE AND ADULT.
CC   -!- SIMILARITY: BELONGS TO THE URICASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X51940; CAA36203.1; -.
DR   PIR; S08676; S08676.
DR   PIR; A35920; A35920.
DR   FLYBASE; FBgn0003961; Uro.
DR   PFAM; PF01014; Uricase; 1.
DR   PRINTS; PR00093; URICASE.
DR   PROSITE; PS00342; MICROBODIES_CTER; 1.
DR   PROSITE; PS00366; URICASE; 1.
KW   Oxidoreductase; Peroxisome; Purine metabolism.
FT   SITE        350    352       MICROBODY TARGETING SIGNAL (POTENTIAL).
SQ   SEQUENCE   352 AA;  40034 MW;  C5E3D44E76627865 CRC64;

  Query Match           32.4%;  Score 81;  DB 1;  Length 352;
  Best Local Similarity  27.6%;  Pred. No. 3.08e-01;
  Matches    8;  Conservative   12;  Mismatches   8;  Indels   1;  Gaps   1;

Db     277 FSVQHTLYL-SERQVLDVLPQVSVISMIM 304
           |::  ||: ::|  | :| |::: :|::
Qy       3 PAIPTQLYFPTDRSFLAILHQLTLTTMSL 31


RESULT    3
ID   DHB1_HUMAN     STANDARD;      PRT;   327 AA.
AC   P14061;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   ESTRADIOL 17 BETA-DEHYDROGENASE 1 (EC 1.1.1.62) (17-BETA-HSD 1)
DE   (PLACENTAL 17 BETA-HYDROXYSTEROID DEHYDROGENASE) (20 ALPHA-
DE   HYDROXYSTEROID DEHYDROGENASE) (20-ALPHA-HSD) (E2DH).
GN   HSD17B1 OR EDH17B1 OR EDHB17 OR E17KSR.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RX   MEDLINE; 89384667.
RA   Luu-The V., Labrie C., Zhao H.F., Couet J., Lachance Y., Simard J.,
RA   Leblanc G., Labrie F.;
RT   "Characterization of cDNAs for human estradiol 17 beta-dehydrogenase
RT   and assignment of the gene to chromosome 17: evidence of two mRNA
RT   species with distinct 5'-termini in human placenta.";
RL   Mol. Endocrinol. 3:1301-1309(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90231840.
RA   Luu-The V., Labrie C., Simard J., Lachance Y., Zhao H.F., Couet J.,
RA   Leblanc G., Labrie F.;
RT   "Structure of two in tandem human 17 beta-hydroxysteroid
RT   dehydrogenase genes.";
RL   Mol. Endocrinol. 4:268-275(1990).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90328695.
RA   Luu-The V., Labrie C., Zhao H.F., Couet J., Lachance Y., Simard J.,
RA   Cote J., Leblanc G., Lagace L., Berube D., Gagne R., Labrie F.;
RT   "Purification, cloning, complementary DNA structure, and predicted
RT   amino acid sequence of human estradiol 17 beta-dehydrogenase.";
RL   Ann. N.Y. Acad. Sci. 595:40-52(1990).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RX   MEDLINE; 89031223.
RA   Peltoketo H., Isomaa V., Maentausta O., Vihko R.;
RT   "Complete amino acid sequence of human placental 17 beta-
RT   hydroxysteroid dehydrogenase deduced from cDNA.";
RL   FEBS Lett. 239:73-77(1988).
RN   [5]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93011163.
RA   Peltoketo H.E., Isomma V., Vihko R.;
RT   "Genomic organization and DNA sequences of human 17 beta-
RT   hydroxysteroid dehydrogenase genes and flanking regions: localization
RT   of multiple Alu sequences and putative cis-acting elements.";
RL   Eur. J. Biochem. 209:459-466(1992).
RN   [6]
RP   SEQUENCE FROM N.A.
RA   Shen Y., Gibbs R.A.;
RL   Submitted (SEP 1995) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE OF 1-23.
RX   MEDLINE; 72238162.
RA   Burns D.J.W., Engel L.L., Bethune J.L.;
RT   "Amino acid composition and subunit structure. Human placental 17 -
RT   estradiol dehydrogenase.";
RL   Biochemistry 11:2699-2703(1972).
RN   [8]
RP   SEQUENCE OF 51-67.
```

```
RX   MEDLINE; 73221216.
RA   Nicholas J.C., Harris J.I.;
RT   "Human placental 17 oestradiol dehydrogenase. Sequence of a tryptic
RT   peptide containing an essential cysteine.";
RL   FEBS Lett. 29:173-176(1973).
RN   [9]
RP   SEQUENCE OF 204-224.
RX   MEDLINE; 86159759.
RA   Murdock G.L., Chin C.C., Warren J.C.;
RT   "Elucidation of the chemical nature of the steady-state intermediates
RT   in the mechanism of carboxypeptidase A.";
RL   Biochemistry 25:646-651(1986).
RN   [10]
RP   SEQUENCE OF 219-224.
RX   MEDLINE; 84008135.
RA   Murdock G.L., Chin C.C., Offord R.E., Bradshaw R.A., Warren J.C.;
RT   "Human placental estradiol 17 beta dehydrogenase. Identification of a
RT   single histidine residue affinity-labeled by both 3-
RT   bromoacetoxyestrone and 12 beta bromoacetoxy-4 estriene 4,17 dione.";
RL   J. Biol. Chem. 258:11460-11464(1983).
RN   [11]
RP   SIMILARITY TO SHORT CHAIN DEHYDROGENASES.
RX   MEDLINE; 89330506.
RA   Baker M.E.;
RT   "Human placental 17 beta-hydroxysteroid dehydrogenase is homologous
RT   to NodG protein of Rhizobium meliloti.";
RL   Mol. Endocrinol. 3:881-884(1989).
RN   [12]
RP   X-RAY CRYSTALLOGRAPHY (2.2 AND 1.7 ANGSTROMS).
RX   MEDLINE; 95393021.
RA   Ghosh D., Pletnev V.Z., Zhu D.W., Wawrzak Z., Duax W.L., Pangborn W.,
RA   Labrie F., Lin S.-X.;
RT   "Structure of human estrogenic 17 beta-hydroxysteroid dehydrogenase
RT   at 2.20-A resolution.";
RL   Structure 3:503-513(1995).
RN   [13]
RP   X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).
RX   MEDLINE; 96398680.
RA   Breton R., Housset D., Mazza C., Fontecilla-Camps J.-C.;
RT   "The structure of a complex of human 17beta-hydroxysteroid
RT   dehydrogenase with estradiol and NADP+ identifies two principal
RT   targets for the design of inhibitors.";
RL   Structure 4:905-915(1996).
RN   [14]
RP   X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).
RX   MEDLINE; 98192605.
RA   Mazza C., Breton R., Housset D., Fontecilla-Camps J.-C.;
RT   "Unusual charge stabilization of NADP+ in 17beta-hydroxysteroid
RT   dehydrogenase.";
RL   J. Biol. Chem. 273:8145-8152(1998).
RN   [15]
RP   VARIANTS VAL-237 AND GLY-312.
RX   MEDLINE; 93278403.
RA   Normand T., Narod S., Labrie F., Simard J.;
RT   "Detection of polymorphisms in the estradiol 17 beta-hydroxysteroid
RT   dehydrogenase II gene at the EDH17B2 locus on 17q11 q21.";
RL   Hum. Mol. Genet. 2:479-483(1993).
CC   -!- FUNCTION: FAVORS THE REDUCTION OF ESTROGENS AND ANDROGENS.
CC       ALSO HAS 20-ALPHA HSD ACTIVITY. USES PREFERENTIALLY NADH.
CC   -!- CATALYTIC ACTIVITY: ESTRADIOL 17 BETA + NAD(+) = ESTRONE +
CC       NADH.
CC   -!- PATHWAY: BIOSYNTHESIS OF ESTROGENS.
CC   -!- SUBUNIT: HOMODIMER.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASE/REDUCTASE
CC       (SDR) FAMILY
CC   ---------- ---------- ----------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ---------- ----------
DR   EMBL; M27148; AAB16941.1;
DR   EMBL; X13440; CAA31792.1;
DR   EMBL; M36263; AAA35600.1;
DR   EMBL; M84472; AAB16942.1;
DR   EMBL; U34879; AAD05019.1;
DR   PIR; A36081; DEHUE7.
DR   PDB; 1EDS; 12-FEB-97.
DR   PDB; 1EDF; 12-FEB-97.
DR   PDB; 1EDU; 27-MAY-98.
DR   PDB; 1EDV; 27-MAY-98.
DR   PDB; 1EDW; 27-MAY-98.
DR   PDB; 1BHS; 07-DEC-96.
DR   PDB; 1IOL; 07-JUL-97.
DR   PDB; 1A27; 27-MAY-98.
DR   PDB; 1EQU; 02-FEB-99.
DR   MIM; 109684; -.
DR   PFAM; PF00106; adh_short; 1.
DR   PROSITE; PS00061; ADH_SHORT; 1.
KW   Steroid biosynthesis; Oxidoreductase; NAD; Multigene
KW   Polymorphism; 3D-structure.
FT   INIT_MET      0      0
FT   NP_BIND       2     21       NAD.
FT   ACT_SITE    142    142       SUBSTRATE BINDING.
FT   ACT_SITE    155    155
FT   DOMAIN       90    ...       HYDROPHOBIC (MICRO-
FT                                 STEROID BINDING).
FT   VARIANT     237    237       A -> V.
FT                                 /FTId=VAR_006951.
FT   VARIANT     312    312       S -> G.
FT                                 /FTId=VAR_006952.
FT   CONFLICT      2      2       P -> E (IN REF. 7).
SQ   SEQUENCE   327 AA;  34849 MW;  B24102D420371992 CR

  Query Match          42.0%;  Score 80;  DB 1;  Length
  Best Local Similarity  4%.6%;  Pred. No. 4.39e-
  Matches   10;  Conservative    ;  Mismatches   4;

Db     244 LKAPKPTLRYFTTER FLPLIE 264
           :|:: || || |:| ::::::
Qy       1 MRPAIPTQLYFPTDRSFLAIIE ..

RESULT    4
ID   ANT3_MOUSE     STANDARD;      PRT;   465 AA.
AC   P32261;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   ANTITHROMBIN-III PRECURSOR (ATIII).
GN   AT3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94069082.
RA   Wu J.K., Sheffield W.P., Perlmutter M.A.;
RT   "Molecular cloning and cell-free expression of mou
RT   III.";
RL   Thromb. Haemost. 68:291-296(1992).
CC   -!- FUNCTION: MOST IMPORTANT SERINE PROTEASE INHIB
CC       THAT REGULATES THE BLOOD COAGULATION CASCADE. A
CC       THROMBIN AS WELL AS FACTORS IXA, XA AND XIA. I
CC       ACTIVITY IS GREATLY ENHANCED IN THE PRESENCE O
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- TISSUE SPECIFICITY: PLASMA.
CC   -!- SIMILARITY: BELONGS TO THE SERPIN FAMILY.
CC   ---------- ---------- ----------
CC   This SWISS-PROT entry is copyright. It is produced
CC   between the Swiss Institute of Bioinformatics and
CC   the European Bioinformatics Institute. There are
```

```
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; S47225; AAB23965.1; -.
DR   HSSP; P01008; 1ATH.
DR   MGD; MGI:88095; AT3.
DR   PFAM; PF00079; serpin; 1.
DR   PROSITE; PS00284; SERPIN; 1.
KW   Serpin; Serine protease inhibitor; Glycoprotein; Heparin-binding;
KW   Plasma; Blood coagulation; Signal.
FT   SIGNAL       1     32       BY SIMILARITY.
FT   CHAIN       33    465       ANTITHROMBIN III.
FT   BINDING     82     82       HEPARIN (BY SIMILARITY).
FT   BINDING    162    162       HEPARIN (BY SIMILARITY).
FT   BINDING    178    178       HEPARIN (BY SIMILARITY).
FT   ACT_SITE   426    427       REACTIVE BOND.
FT   DISULFID    41    161       BY SIMILARITY.
FT   DISULFID    54    128       BY SIMILARITY.
FT   DISULFID   280    463       BY SIMILARITY.
FT   CARBOHYD   129    129       POTENTIAL.
FT   CARBOHYD   168    168       POTENTIAL.
FT   CARBOHYD   188    188       POTENTIAL.
FT   CARBOHYD   225    225       POTENTIAL.
SQ   SEQUENCE   465 AA;  52003 MW;  5CE087E98874E35D CRC64;

  Query Match           42.0%;  Score 80;  DB 1;  Length 465;
  Best Local Similarity  24.2%;  Pred. No. 4.39e-01;
  Matches    8;  Conservative   16;  Mismatches   9;  Indels   0;  Gaps   0;

Db    426 RSLNPNRVTPKANEPFLVLIREVALNTIPMCR 458
          |:  | :: | :::|:|| :::::::| |: :: :
Qy      2 RPAIPIQLYFPTDRSFLAILHQLTLITMSLVDQ 34


RESULT    5
 ID   NU1M_DASNO     STANDARD;      PRT;   318 AA.
AC   O21325;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 1 (EC 1.6.5.3).
GN   MTND1 OR ND1 OR NADH1.
OS   Dasypus novemcinctus (Nine banded armadillo).
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Edentata; Dasypodidae; Dasypus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97357423.
RA   Arnason U., Gullberg A., Janke A.;
RT   "Phylogenetic analyses of mitochondrial DNA suggest a sister group
RT   relationship between Xenarthra (Edentata) and Ferungulates.";
RL   Mol. Biol. Evol. 14:762-768(1997).
CC   -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
CC   -!- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; Y11832; CAA72529.1; -.
DR   PFAM; PF00146; NADHdh; 1.
DR   PROSITE; PS00667; COMPLEX1_ND1_1; 1.
DR   PROSITE; PS00668; COMPLEX1_ND1_2; 1.
KW   Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.
SQ   SEQUENCE   318 AA;  35591 MW;  26E2DB15B712FF6A CRC64;

  Query Match           30.8%;  Score 77;  DB 1;  Length 318;
  Best Local Similarity  30.3%;  Pred. No. 1.24e+00;
  Matches   23;  Conservative   19;  Mismatches   12;  Indels   1;  Gaps   1;

Db    233 MMNALTAILFFGAMHNPMFPELHTLNLITKTLI 265
          |  |: : |:| :   : :: || | | |::|:
Qy      1 MRPAIPTQLYFPT-DRSFLAILHQLTLTTMSLV 32


RESULT    6
ID   BCPA_PROAE     STANDARD;      PRT;   366 AA.
AC   P11741;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   BACTERIOCHLOROPHYLL A PROTEIN (BCHL A PROTEIN) (BCP).
OS   Prosthecochloris aestuarii.
OC   Bacteria; Green sulfur bacteria; Prosthecochloris.
RN   [1]
RP   SEQUENCE.
RC   STRAIN=2K;
RX   MEDLINE; 86140151.
RA   Daurat-Larroque S.T., Brew K., Fenna R.E.;
RT   "The complete amino acid sequence of a bacteriochlorophyll a-protein
RT   from Prosthecochloris aestuarii.";
RL   J. Biol. Chem. 261:3607-3615(1986).
RN   [2]
RP   X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS), AND PARTIAL SEQUENCE.
RX   MEDLINE; 86281701.
RA   Tronrud D.E., Schmid M.F., Matthews B.W.;
RT   "Structure and X-ray amino acid sequence of a bacteriochlorophyll a
RT   protein from Prosthecochloris aestuarii refined at 1.9-A
RT   resolution.";
RL   J. Mol. Biol. 188:443-454(1986).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
RX   MEDLINE; 80029713.
RA   Matthews B.W., Fenna R.E., Bolognesi M.C., Schmid M.F., Olson J.M.;
RT   "Structure of a bacteriochlorophyll a-protein from the green
RT   photosynthetic bacterium Prosthecochloris aestuarii.";
RL   J. Mol. Biol. 131:259-285(1979).
CC   -!- FUNCTION: INTERMEDIARY IN THE TRANSFER OF EXCITATION ENERGY FROM
CC       THE CHLOROPHYLL TO THE REACTION CENTERS.
CC   -!- SUBUNIT: HOMOTRIMER.  EACH SUBUNIT CONTAINS 7 MOLECULES OF
CC       BACTERIOCHLOROPHYLL A.
DR   PIR; A25330; A25330.
DR   PIR; A25302; A25302.
DR   PDB; 4BCL; 15-JUL-98.
KW   Electron transport; Photosynthesis; Reaction center; Magnesium;
KW   3D-structure; Bacteriochlorophyll.
FT   BINDING     110    110       BACTERIOCHLOROPHYLL A, 1.
FT   BINDING     145    145       BACTERIOCHLOROPHYLL A, 6.
FT   BINDING     290    290       BACTERIOCHLOROPHYLL A, 4.
FT   BINDING     297    297       BACTERIOCHLOROPHYLL A, 7.
FT   BINDING     298    298       BACTERIOCHLOROPHYLL A, 3.
FT   STRAND       10     19
FT   TURN         22     23
FT   STRAND       27     35
FT   STRAND       46     56
FT   STRAND       64     74
FT   TURN         75     76
FT   STRAND       77     89
FT   STRAND       94    105
FT   TURN        106    107
FT   STRAND      108    122
FT   HELIX       127    132
FT   STRAND      141    152
FT   HELIX       156    168
FT   HELIX       176    184
FT   TURN        186    188
FT   HELIX       189    195
FT   TURN        196    197
FT   STRAND      198    200
```

```
FT   STRAND       204     210
FT   STRAND       218     229
FT   HELIX        234     237
FT   TURN         238     242
FT   HELIX        245     247
FT   STRAND       252     260
FT   STRAND       265     273
FT   TURN         274     275
FT   STRAND       277     281
FT   TURN         282     283
FT   STRAND       284     287
FT   HELIX        292     298
FT   TURN         299     303
FT   STRAND       308     316
FT   TURN         319     320
FT   STRAND       322     327
FT   STRAND       332     334
FT   TURN         335     336
FT   STRAND       337     341
FT   HELIX        343     353
FT   TURN         354     355
FT   STRAND       361     365
SQ   SEQUENCE   466 AA;  40276 MW;  57242734E53615B1 CRC64;

  Query Match           30.8%;  Score 77;  DB 1;  Length 466;
  Best Local Similarity  36.4%;  Pred. No. 1.24e-00;
  Matches   12;  Conservative    3;  Mismatches  17;  Indels   1;  Gaps   1;

Db    129 VRRNIPNPIYMQ-CRQFHDILMKVPLDNNDLVD 160
          :|  ||  :|   | |  ||  : |    |||
Qy      1 MRPAIPIQLYFPIDRSFLAILHQLILTTMSLVD 33


RESULT    7
ID   ANT3_HUMAN      STANDARD;      PRT:   464 AA.
AC   P01008;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   ANTITHROMBIN-III PRECURSOR (ATIII).
GN   AT3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mateleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 83134280.
RA   Bock S.C., Wion K.L., Vehar G.A., Lawn R.M.;
RT   "Cloning and expression of the cDNA for human antithrombin III.";
RL   Nucleic Acids Res. 10:8113-8125(1982).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 83169777.
RA   Chandra T., Stackhouse R., Kidd V.J., Woo S.L.C.;
RT   "Isolation and sequence characterization of a cDNA clone of human
RT   antithrombin III.";
RL   Proc. Natl. Acad. Sci. U.S.A. 80:1845-1848(1983).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93237227.
RA   Olds R.J., Lane D.A., Chowdhury V., de Stefano V., Leone G.,
RA   Thein S.L.;
RT   "Complete nucleotide sequence of the antithrombin gene: evidence for
RT   homologous recombination causing thrombophilia.";
RL   Biochemistry 32:4216-4224(1993).
RN   [4]
RP   SEQUENCE OF 42-464 FROM N.A.
RX   MEDLINE; 83238456.
RA   Prochownik E.V., Markham A.F., Orkin S.H.;
RT   "Isolation of a cDNA clone for human antithrombin III.";
RL   J. Biol. Chem. 258:8389-8394(1983).
RN   [5]
RP   SEQUENCE OF 33-464, CARBOHYDRATE-BINDING SITES, AND DISULFIDE BONDS.
RA   Petersen T.E., Dudek Wojciechowska G., Sottrup Je
RA   Magnusson S.;
RT   "Primary structure of antithrombin-III (heparin c
RT   homology between alpha 1-antitrypsin and antithror
RL   (In) Collen D., Wiman B., Verstraete M. (eds.):
RL   The physiological inhibitors of blood coagulation
RL   pp.43-54, Elsevier, Amsterdam (1979).
RN   [6]
RP   ACTIVE SITE.
RX   MEDLINE; 81212814.
RA   Bjoerk I., Danielsson A., Fenton J.W. II, Joernva
RT   "The site in human antithrombin for functional pr
RL   by human thrombin.";
RL   FEBS Lett. 126:257-260 (1981).
RN   [7]
RP   HEPARIN-BINDING SITE.
RX   MEDLINE; 84111578.
RA   Blackburn M.N., Smith R.L., Carson J., Sibley C.
RT   "The heparin-binding site of antithrombin III. Id
RT   critical tryptophan in the amino acid sequence.";
RL   J. Biol. Chem. 259:9934-9942(1984).
RN   [8]
RP   MUTAGENESIS OF ALA 414.
RX   MEDLINE; 91192143.
RA   Austin K.C., Rachubinski R.A., Blackburn M.N.;
RT   "Site-directed mutagenesis of alanine-382 of human
RL   FEBS Lett. 280:254-258(1991).
RN   [9]
RP   X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
RX   MEDLINE; 94373498.
RA   Carrell R.W., Stein P.E., Fermi G., Wardell M.R.;
RT   "Biological implications of a 3 A structure of di
RL   Structure 2:257-270(1994).
RN   [10]
RP   X-RAY CRYSTALLOGRAPHY (3.2 ANGSTROMS).
RX   MEDLINE; 95384753.
RA   Schreuder H.A., de Boer B., Dijkema R., Mulders J.
RA   Grootenhuis P.D.J., Hol W.G.J.;
RT   "The intact and cleaved human antithrombin III com
RT   serpin-proteinase interactions.";
RL   Nat. Struct. Biol. 1:48-54(1994).
RN   [11]
RP   X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
RX   MEDLINE; 97220238.
RA   Skinner R., Abrahams J. P., Whisstock J.C., Lesk A
RA   Wardell M.R.;
RT   "The 2.6 A structure of antithrombin indicates a
RT   change at the heparin binding site.";
RL   J. Mol. Biol. 266:601-609(1997)
RN   [12]
RP   X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS).
RX   MEDLINE; 98437366.
RA   Skinner R., Chang W.S.W., Jin L., Pei X.Y., Huntin
RA   Abrahams J.-P., Carrell R.W., Lomas D.A.;
RT   "Implications for function and therapy of a 2.9 A
RT   binary-complexed antithrombin.";
RL   J. Mol. Biol. 283:9-14(1998).
RN   [13]
RP   REVIEW.
RX   MEDLINE; 91129302.
RA   Mourey L., Samama J. P., Delarue M., Choay J., Lor
RA   Petitou M., Moras D.;
RT   "Antithrombin III: structural and functional aspec
RL   Biochimie 72:599-608(1990).
RN   [14]
RP   REVIEW ON VARIANTS.
RX   MEDLINE; 94054329.
RA   Lane D.A., Olds R.J., Boisclair M., Chowdhury V.,
RA   Cooper D.N., Blajchman M., Perry D., Emmerich J., A
RT   "Antithrombin III mutation database: first update.
RT   and its Inhibitors Subcommittee of the Scientific
RT   Committee of the International Society on Thrombos
RT   Haemostasis.";
```

```
RL   Thromb. Haemost. 70:361-369(1993).
RN   [15]
RP   REVIEW ON VARIANTS.
RX   MEDLINE; 95269065.
RA   Stein P.E., Carrell R.W.;
RT   "What do dysfunctional serpins tell us about molecular mobility and
RT   disease?";
RL   Nat. Struct. Biol. 2:96-113(1995).
RN   [16]
RP   REVIEW ON VARIANTS.
RX   MEDLINE; 96263733.
RA   Perry D.J., Carrell R.W.;
RT   "Molecular genetics of human antithrombin deficiency.";
RL   Hum. Mutat. 7:7-22(1996).
RN   [17]
RP   VARIANT UTAH.
RX   MEDLINE; 89050967.
RA   Bock S.C., Marrinan J.A., Radziejewska E.;
RT   "Antithrombin III Utah: proline-407 to leucine mutation in a highly
RT   conserved region near the inhibitor reactive site.";
RL   Biochemistry 27:6171-6178(1988).
RN   [18]
RP   VARIANT TOYAMA.
RX   MEDLINE; 84119472.
RA   Koide T., Odani S., Takahashi K., Ono T., Sakuragawa N.;
RT   "Antithrombin III Toyama: replacement of arginine-47 by cysteine in
RT   hereditary abnormal antithrombin III that lacks heparin-binding
RT   ability.";
RL   Proc. Natl. Acad. Sci. U.S.A. 81:289-293(1984).
RN   [19]
RP   VARIANT BASEL.
RX   MEDLINE; 86111754.
RA   Chang J.Y., Tran T.H.;
RT   "Antithrombin III Basel. Identification of a Pro-Leu substitution in
RT   a hereditary abnormal antithrombin with impaired heparin cofactor
RT   activity.";
RL   J. Biol. Chem. 261:1174-1176(1986).
RN   [20]
RP   VARIANT DENVER.
RX   MEDLINE; 87109210.
RA   Stephens A.W., Thalley B.S., Hirs C.H.W.;
RT   "Antithrombin-III Denver, a reactive site variant.";
RL   J. Biol. Chem. 262:1044-1048(1987).
RN   [21]
RP   VARIANT HAMILTON.
RX   MEDLINE; 89027076.
RA   Devrak-Kizuk R., Chui D.H.K., Prochownik E.V., Carter C.J.,
RA   Ofosu F.A., Blajchman M.A.;
RT   "Antithrombin-III-Hamilton: a gene with a point mutation (guanine to
RT   adenine) in codon 382 causing impaired serine protease reactivity.";
RL   Blood 72:1518-1523(1988).
RN   [22]
RP   VARIANTS GLASGOW AND NORTHWICK PARK.
RX   MEDLINE; 88186869.
RA   Erdjument H., Laned D.A., Panico M., di Marzo V., Morris H.R.;
RT   "Single amino acid substitutions in the reactive site of antithrombin
RT   leading to thrombosis. Congenital substitution of arginine 393 to
RT   cysteine in antithrombin Northwick Park and to histidine in
RT   antithrombin Glasgow.";
RL   J. Biol. Chem. 263:5589-5593(1988).
RN   [23]
RP   VARIANT CHICAGO.
RX   MEDLINE; 89388698.
RA   Erdjument H., Lane D.A., Panico M., di Marzo V., Morris H.R.,
RA   Bauer K., Rosenberg R.D.;
RT   "Antithrombin Chicago, amino acid substitution of arginine 393 to
RT   histidine.";
RL   Thromb. Res. 54:613-619(1989).
RN   [24]
RP   VARIANT ROUEN-4.
RX   MEDLINE; 90306344.
RA   Borg J.Y., Brennan S.O., Carrell R.W., George P., Perry D.J., Shaw J.;
RT   "Antithrombin Rouen-IV 24 Arg-->Cys. The amino-terminal contribution
RT   to heparin binding.";
RL   FEBS Lett. 266:163-166(1990).
RN   [25]
RP   VARIANT DUBLIN.
RX   MEDLINE; 91032211.
RA   Daly M., Bruce D., Perry D.J., Price J., Harper P.L., O'Meara A.,
RA   Carrell R.W.;
RT   "Antithrombin Dublin (-3 Val-->Glu): an N terminal variant which has
RT   an aberrant signal peptidase cleavage site.";
RL   FEBS Lett. 273:87-90(1990).
RN   [26]
RP   VARIANT GENEVA.
RX   MEDLINE; 91035423.
RA   Gandrille S., Aiach M., Lane D.A., Vidaud D., Molho-Sabatier P.,
RA   Caso R., de Moerloose P., Fiessinger J.-N., Clauser E.;
...
Note: remainder of annotations omitted.

  Query Match          30.8%;  Score 77;  DB 1;  Length 464;
  Best Local Similarity  24.2%;  Pred. No. 1.24e+00;
  Matches    8;  Conservative   15;  Mismatches  10;  Indels   0;  Gaps   0;

Db    425 RSLNPNPVTFKANPPFLVFIPEVPLNTIIPMGR 457
          |:  | :: | ::|:|| ::::: | |: :: :
Qy      2 KPAIFTGLYTFTDKSFLAILHQETLTTMGLVDQ 34


RESULT    8
ID   MCM2_DROME      STANDARD;       PRT;   887 AA.
AC   P49735;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   DNA REPLICATION LICENSING FACTOR MCM2.
GN   MCM2.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EYE ANTENNAL DISK;
RX   MEDLINE; 95347580.
RA   Treisman J.E., Follette P.J., O'Farrell P.H., Rubin G.M.;
RT   "Cell proliferation and DNA replication defects in a Drosophila MCM2
RT   mutant.";
RL   Genes Dev. 9:1709-1715(1995).
CC   -!- FUNCTION: ACT AS A FACTOR THAT LICENSE THE DNA FOR ONE AND ONLY
CC       ONE ROUND OF REPLICATION PER CELL CYCLE. REQUIRED FOR DNA
CC       REPLICATION AND CELL PROLIFERATION.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC   -!- SIMILARITY: BELONGS TO THE MCM FAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; L42762; AAB86617.1; -.
DR   FLYBASE; FBgn0014861; Mcm2.
DR   PFAM; PF00493; MCM; 1.
DR   PROSITE; PS00847; MCM_1; 1.
DR   PROSITE; PS50051; MCM_2; 1.
KW   Transcription regulation; DNA-binding; Nuclear protein;
KW   DNA replication; Zinc-finger, ATP binding; Cell cycle.
FT   ZN_FING     314    340       C4-TYPE (POTENTIAL).
FT   DOMAIN      458    665       MCM.
FT   NP_BIND     508    515       ATP (POTENTIAL).
SQ   SEQUENCE   887 AA;  100414 MW;  26A7092109F09CBE CRC64;
```

```
  Query Match             30.8%;  Score 77;  DB 1;  Length 887;
  Best Local Similarity   38.2%;  Pred. No. 1.24e+00;
  Matches   13;  Conservative    9;  Mismatches   10;  Indels    2;  Gaps    2;

Db     791 MRSTPQKYLSPQKDHSELLFPFTLRQLLLDQLAYI 824
           ||:::   | :  :|| :|  ||:|||  :: :
Qy       1 MRPAIPTQLYFPTDRS-PLA-IIHQLTITTMSLV 32


RESULT    9
ID   URIC_DROVI     STANDARD;      PRT;   342 AA.
AC   P23104;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   URICASE (EC 1.7.3.3) (URATE OXIDASE)
GN   URO OR UO.
OS   Drosophila virilis (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wallrath L.L.;
RL   Submitted (JAN-1991) to the EMBL/GenBank/DDBJ databases
CC   -!- CATALYTIC ACTIVITY: URATE + O(2) + H(2)O = 5-HYDROXYISOURATE +
CC       H(2)O(2) (5-HYDROXYISOURATE DECOMPOSE TO FORM ALLANTOIN).
CC   -!- ENZYME REGULATION: REPRESSED BY 20-HYDROXYECDYSONE.
CC   -!- SUBCELLULAR LOCATION: PEROXISOMAL.
CC   -!- TISSUE SPECIFICITY: MALPIGHIAN TUBULES.
CC   -!- DEVELOPMENTAL STAGE: THIRD INSTAR LARVAE AND ADULT.
CC   -!- SIMILARITY: BELONGS TO THE URICASE FAMILY.
CC   ---------------------------------------- ----------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------- ----------------------------------------
DR   EMBL; X57114; CAA40497.1; -.
DR   PIR; S29133; S29133.
DR   FLYBASE; FBgn0013101; Dvir\Uro.
DR   PFAM; PF01014; Uricase; 1.
DR   PRINTS; PR00093; URICASE.
DR   PROSITE; PS00342; MICROBODIES_CTER; 1.
DR   PROSITE; PS00366; URICASE; 1.
KW   Oxidoreductase; Peroxisome; Purine metabolism.
FT   SITE        340    342       MICROBODY TARGETING SIGNAL (POTENTIAL).
SQ   SEQUENCE   342 AA;  38724 MW;  4DE10B5485294723 CRC64;

  Query Match             30.4%;  Score 76;  DB 1;  Length 342;
  Best Local Similarity   24.1%;  Pred. No. 1.75e+00;
  Matches    7;  Conservative   13;  Mismatches    8;  Indels    1;  Gaps    1;

Db     267 PSVQHTLYL-SEKQVLDVLPQVSVVSMFM 294
           |::   ||: ::: : :| |::: :|::
Qy       3 PAIPTQLYFPTDRSFLAIHQLILIIMSL 31


RESULT   10
ID   CYB_SORCA      STANDARD;      PRT;   379 AA.
AC   Q79450; Q21737;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   CYTOCHROME B.
GN   MTCYB OR COB OR CYTB.
OS   Sorex caecutiens (Laxmann's shrew).
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Insectivora; Soricidae; Sorex.
RN   [1]
RP   SEQUENCE OF 1-134 FROM N.A.
RC   STRAIN=665, SIZAF1, SIZAF2, AND S25/95; TISSUE=LIV
RA   Ohdachi S., Masuda R., Abe H., Adachi J., Dokuchae
RA   Haukisalmi V., Yoshida M.;
RT   "Molecular phylogeny from cytochrome sequences :
RT   cytochrome b gene and evolutionary history of Eur
RT   (Mammalia, Insectivora).";
RL   Zool. Sci. 14:527-532(1997).
RN   [2]
RP   SEQUENCE OF 44-379 FROM N.A.
RA   Fumagalli L., Taberlet P., Stewart D.T., Gielly L
RA   Vogel P.;
RT   "Molecular phylogeny and evolution of Sorex shrews
RT   Insectivora) inferred from mitochondrial DNA seque
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ data
CC   -!- FUNCTION: COMPONENT OF THE UBIQUINOL CYTOCHROM
CC       COMPLEX (COMPLEX III OR CYTOCHROME B C1 COMPL
CC       RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHE
CC       COUPLED TO ATP SYNTHESIS.
CC   -!- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHIC
CC       BOUND TO THE PROTEIN.
CC   -!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 AR
CC       CYTOCHROME C1 AND THE RIESKE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY
CC   ---------- ----------  ----------
CC   This SWISS-PROT entry is copyright. It is produced
CC   between the Swiss Institute of Bioinformatics an
CC   the European Bioinformatics Institute. There are
CC   use by non-profit institutions as long as its
CC   modified and this statement is not removed. Usage
CC   entities requires a license agreement (See http://
CC   or send an email to license@isb-sib.ch).
CC   ---------- ----------  ----------
DR   EMBL; D85347; BAA21340.1; -.
DR   EMBL; D85349; BAA21342.1; -.
DR   EMBL; D85360; BAA21353.1; -.
DR   EMBL; D85367; BAA21360.1; -.
DR   EMBL; A70203; CAA04077.1; -.
DR   PFAM; PF00032; cytochrome_b_C; 1.
DR   PFAM; PF00033; cytochrome_b_N; 1.
DR   PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
DR   PROSITE; PS00193; CYTOCHROME_B_QO; 1.
KW   Mitochondrion; Electron transport; Respiratory cha
KW   Heme.
FT   METAL        83     83       IRON 1 (HEME B562 AX
FT   METAL        97     97       IRON 2 (HEME B566 AX
FT   METAL       182    182       IRON 2 (HEME B562 AX
FT   METAL       196    196       IRON 1 (HEME B566 AX
SQ   SEQUENCE   379 AA;  42542 MW;  EA22AD90B21AA59E1 CR

  Query Match             30.3%;  Score 75;  DB 1;  Leng
  Best Local Similarity   42.9%;  Pred. No. 2.45e+00;
  Matches    12;  Conservative    6;  Mismatches    8;

Db     218 IPFHPYY-TIKDILGVL LLILVLMSLV 248
           || . |: | : :|::|  |  . |||
Qy       5 IPTQLYFPTDRSFLAIHQLILIIMSLV 32


RESULT   11
ID   CYB_SPERI      STANDARD;      PRT;   379 AA.
AC   P49341;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   CYTOCHROME B.
GN   MTCYB OR COB OR CYTB.
OS   Spermophilus richardsonii (Richardson's ground squ
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata
OC   Mammalia; Eutheria; Rodentia; Sciuromorpha; Sciuri
OC   Spermophilus.
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94326000.
RA   Thomas W.K., Martin S.L.;
RT   "A recent origin of marmots.";
RL   Mol. Phylogenet. Evol. 2:230-236(1993).
CC   -!- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC       COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC       RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC       COUPLED TO ATP SYNTHESIS.
CC   -!- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC       BOUND TO THE PROTEIN.
CC   -!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC       CYTOCHROME C1 AND THE RIESKE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch)
CC   ---------------------------------------------------------------------
DR   EMBL; S73150; AAC60686.1; -.
DR   PFAM; PF00032; cytochrome_b_C; 1.
DR   PFAM; PF00033; cytochrome_b_N; 1.
DR   PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
DR   PROSITE; PS00193; CYTOCHROME_B_QO; 1.
KW   Electron transport; Mitochondrion; Respiratory chain; Transmembrane;
KW   Heme.
FT   METAL        83     83       IRON 1 (HEME B562 AXIAL LIGAND).
FT   METAL        97     97       IRON 2 (HEME B566 AXIAL LIGAND).
FT   METAL       182    182       IRON 2 (HEME B562 AXIAL LIGAND).
FT   METAL       196    196       IRON 1 (HEME B566 AXIAL LIGAND).
SQ   SEQUENCE   379 AA;  42924 MW;  01CFB206E1539D6B CRC64;

  Query Match           30.0%;  Score 75;  DB 1;  Length 379;
  Best Local Similarity  39.3%;  Pred. No. 2.45e+00;
  Matches   11;  Conservative    8;  Mismatches   7;  Indels   2;  Gaps   2;

Db    218 IPFHPYY-TIKDILGVL-LLILALMTLV 243
          || : |: | : :|::|  | |: |:||
Qy      5 IPTQLYFPTDRSFLAILHQLILTTMSLV 32


RESULT   12
ID   CYB_MOSMO      STANDARD;      PRT:   379 AA.
AC   O47583;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   CYTOCHROME B.
GN   MTCYB OR COB OR CYTB.
OS   Moschus moschiferus (Siberian musk deer).
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Cervoidea;
OC   Moschidae; Moschus.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Su B., Wang Y.X., Lan H., Wang W., Zhang Y.P.;
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC       COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC       RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC       COUPLED TO ATP SYNTHESIS.
CC   -!- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC       BOUND TO THE PROTEIN.
CC   -!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC       CYTOCHROME C1 AND THE RIESKE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF026883; AAH94605.1; -.
DR   PFAM; PF00032; cytochrome_b_C; 1.
DR   PFAM; PF00033; cytochrome_b_N; 1.
DR   PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
DR   PROSITE; PS00193; CYTOCHROME_B_QO; 1.
KW   Electron transport; Mitochondrion; Respiratory chain; Transmembrane;
KW   Heme.
FT   METAL        83     83       IRON 1 (HEME B562 AXIAL LIGAND).
FT   METAL        97     97       IRON 2 (HEME B566 AXIAL LIGAND).
FT   METAL       182    182       IRON 2 (HEME B562 AXIAL LIGAND).
FT   METAL       196    196       IRON 1 (HEME B566 AXIAL LIGAND).
SQ   SEQUENCE   379 AA;  42919 MW;  F4717E47024F4DCC CRC64;

  Query Match           30.0%;  Score 75;  DB 1;  Length 379;
  Best Local Similarity  42.9%;  Pred. No. 2.45e+00;
  Matches   12;  Conservative    6;  Mismatches   8;  Indels   2;  Gaps   2;

Db    218 IPFHPYY TIKDILGIL LLILILMALV 243
          || : |: | : :|:|| | | |:||
Qy      5 IPTQLYFPTDRSFLAILHQLILTTMSLV 32


RESULT   13
ID   EAT2_CAEEL     STANDARD;      PRT:   532 AA.
AC   Q21453;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   PUTATIVE SODIUM-DEPENDENT EXCITATORY AMINO ACID TRANSPORTER K08F4.4.
GN   K08F4.4.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN BRISTOL N2;
RA   Hembry C.;
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SODIUM:DICARBOXYLATE SYMPORTER FAMILY
CC       (SDF).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z68879; CAA93084.1; -.
DR   WORMPEP; K08F4.4; CE06152.
DR   PFAM; PF00375; SDF; 1.
DR   PRINTS; PR00173; EDTRNSPORT.
DR   PROSITE; PS00713; NA_DICARBOXYL_SYMP_1; 1.
DR   PROSITE; PS00714; NA_DICARBOXYL_SYMP_2; 1.
KW   Hypothetical protein; Transport; Transmembrane; Glycoprotein; Symport.
FT   DOMAIN       1      5       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     6     26       POTENTIAL.
FT   TRANSMEM    46     66       POTENTIAL.
FT   TRANSMEM    83    103       POTENTIAL.
FT   DOMAIN      104    181       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    182    202       POTENTIAL.
FT   TRANSMEM    222    242       POTENTIAL.
FT   TRANSMEM    264    284       POTENTIAL.
```

```
FT   TRANSMEM     352    372       POTENTIAL.
FT   TRANSMEM     383    402       POTENTIAL.
FT   CARBOHYD     164    164       POTENTIAL.
FT   CARBOHYD     169    169       POTENTIAL.
SQ   SEQUENCE   532 AA;  58469 MW;  FA370EAD3796E5DF CRC64;

  Query Match           30.0%;  Score 75;  DB 1;  Length 532;
  Best Local Similarity  42.4%;  Pred. No. 2.45e-00;
  Matches   11;  Conservative    4;  Mismatches  11;  Indels   0;  Gaps   0;

Db    347 TALYFAVAVIFIAQIHNIKLSLMDLV 372
          | ||:    |:| || : |: | ||
Qy      7 IQLYFPTDRSFLAILHQLTLTTMSLV 32


RESULT   14
ID   RPIA_CHLTR     STANDARD;      PRT;   242 AA.
AC   O84215;
DT   15-FEB-2000 (Rel. 39, Created)
DT   15-FEB-2000 (Rel. 39, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   RIBOSE 5-PHOSPHATE ISOMERASE A (EC 5.3.1.6) (PHOSPHORIBOISOMERASE A).
GN   RPIA OR CT213.
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN   [1]
RP   SEQUENCE FROM N.A.
RP   STRAIN D/UW-3/CX.
RX   MEDLINE; 99000809.
RA   Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA   Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA   Davis R.W.;
RT   "Genome sequence of an obligate intracellular pathogen of humans:
RT   Chlamydia trachomatis.";
RL   Science 282:754-759(1998).
CC   -!- CATALYTIC ACTIVITY: D-RIBOSE 5-PHOSPHATE = D-RIBULOSE 5-PHOSPHATE.
CC   -!- PATHWAY: NONOXIDATIVE BRANCH OF THE PENTOSE PHOSPHATE PATHWAY.
CC   -!- SIMILARITY: BELONGS TO THE RIBOSE 5-PHOSPHATE ISOMERASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE001295; AAC67805.1; -.
KW   Isomerase.
SQ   SEQUENCE   242 AA;  26649 MW;  059C38284D2229B2 CRC64;

  Query Match           29.6%;  Score 74;  DB 1;  Length 242;
  Best Local Similarity  34.8%;  Pred. No. 3.43e+00;
  Matches    8;  Conservative   11;  Mismatches   3;  Indels   1;  Gaps   1;

Db      9 FSSDK-FFPTKQKLALEAVALVE 30
          |:::: |::| : |:| :::||:
Qy     11 FPTDRSPLAILHQLTLTTMSLVD 33


RESULT   15
ID   CYB_CRACG      STANDARD;      PRT;   379 AA.
AC   Q34106;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   CYTOCHROME B.
GN   MTCYB OR COB OR CYTB.
OS   Cratogeomys goldmani goldmani (Pappogeomys goldmani goldmani).
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuromorphi; Geomyidae; Cratogeomys.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94184502.
RA   Dewalt T.S., Sudman P.D., Batner M.S., Davis S.K.
RT   "Phylogenetic relationships of pocket gophers (Ge
RT   Pappogeomys) based on mitochondrial DNA cytochrome
RL   Mol. Phylogenet. Evol. 2:193-204(1993).
CC   -!- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROM
CC       COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLE
CC       RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHE
CC       COUPLED TO ATP SYNTHESIS.
CC   -!- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHIC
CC       BOUND TO THE PROTEIN.
CC   -!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX BC1 AR
CC       CYTOCHROME C1 AND THE RIESKE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced
CC   between the Swiss Institute of Bioinformatics an
CC   the European Bioinformatics Institute. There are
CC   use by non-profit institutions as long as its
CC   modified and this statement is not removed. Usage
CC   entities requires a license agreement (See http://
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L11904; AAA97491.1; -.
DR   PFAM; PF00032; cytochrome_b_C; 1.
DR   PFAM; PF00033; cytochrome_b_N; 1.
DR   PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
DR   PROSITE; PS00193; CYTOCHROME_B_QO; 1.
KW   Electron transport; Mitochondrion; Respiratory cha
KW   Heme.
FT   METAL        83     83       IRON 1 (HEME B562 AXI
FT   METAL        97     97       IRON 2 (HEME B566 AXI
FT   METAL       182    182       IRON 2 (HEME B562 AXI
FT   METAL       196    196       IRON 1 (HEME B566 AXI
SQ   SEQUENCE   379 AA;  42952 MW;  7E6D16BC8F64AF1C CR

  Query Match           29.6%;  Score 74;  DB 1;  Length
  Best Local Similarity  35.7%;  Pred. No. 8.43e+00;
  Matches   10;  Conservative    9;  Mismatches   7;  In

Db    218 VPFHPYYIT-KDEMGVILLLIF MIIV 243
          :| : |: | : |:::: :: ::
Qy      5 IPIQLYFPTDRSFLAILHQLTLTTMSLV 32


Search completed: Fri Aug 25 16:59:12 2000
Job time : 17 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 25 18:05:21 2000; MasPar time 8.17 Seconds
288.688 Million cell updates/sec

Tabular output not generated.

Title: >US-09-257-585-7
Description: (1-34) from US09257585.pep
Perfect Score: 250
Sequence: 1 MRPAIPTQLYFPIDRSFLAILHQLILTTMSLVDQ 34

Scoring table: PAM 150
Gap 11

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl12
1:sp_archea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 34.126; Variance 62.025; scale 0.550

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	94	37.6	645	8	O63620	NADH DEHYDROGENASE SUB	1.02e-02
2	91	36.4	622	4	Q9Y6J9	PCAF ASSOCIATED FACTOR	3.03e-02
3	82	32.8	962	14	Q9W6N6	RECOMBINATION AC IVATI	7.13e-01
4	81	32.4	434	5	Q23217	W07G4.2 PROTEIN.	1.00e+00
5	79	31.6	322	1	Q9YEJ5	322AA LONG HYPOTHETICA	1.96e+00
6	79	31.6	476	3	P78998	HYPOTHETICAL 41.5 KD P	1.96e+00
7	78	31.2	459	11	Q9WTI1	ANTITHROMBIN III	2.73e+00
8	77	30.8	144	1	O59576	144AA LONG HYPOTHETICA	3.79e+00
9	77	30.8	379	8	O47997	CYTOCHROME B.	3.79e+00
10	76	30.4	379	8	O36229	CYTOCHROME B.	5.25e+00
11	76	30.4	571	5	O18233	Y57G11C.7 PROTEIN.	5.25e+00
12	76	30.4	962	13	Q9W6N5	RECOMBINATION AC IVATI	5.25e+00
13	76	30.4	1209	5	Q93620	F27D4.6 PROTEIN.	5.25e+00
14	76	30.4	1275	14	O41977	TEGUMENT PROTEIN/PGARA	5.25e+00
15	76	30.4	2609	4	O74390	PUTATIVE TRANSPORT PRO	5.25e+00
16	75	30.0	260	2	Q926F6	POLAR FLAGELLAR ASSEMB	7.26e+00
17	75	30.0	379	8	Q9XP26	CYTOCHROME B.	7.26e+00
18	75	30.0	380	8	O74198	CYTOCHROME B.	7.26e+00
19	75	30.0	406	5	Q22975	COSMID F45F2.	7.26e+00
20	74	29.6	242	2	O84215	RIBOSE 5 P ISOMERASE A	1.00e+01
21	74	29.6	333	1	Q19173	COSMID F07	1.00e+01
22	74	29.6	374	12	Q38993	HYPOTHETICA	1.00e+01
23	74	29.6	379	8	47995	CYTOCHROME	1.00e+01
24	74	29.6	379	8	Q54671	CYTOCHROME	1.00e+01
25	74	29.6	380	8	79221	CYTOCHROME	1.00e+01
26	74	29.6	381	8	Q9XN04	CYTOCHROME	1.00e+01
27	74	29.6	414	1	Q37Y9	HYPOTHETICA	1.00e+01
28	73	29.2	282	2	Q94Z06	HYPOTHETICA	1.37e+01
29	73	29.2	339	8	Q737J1	NADH DEHYDR	1.37e+01
30	73	29.2	347	5	Q87Y2	C50C11.6 P	1.37e+01
31	73	29.2	379	8	Q9XP28	CYTOCHROME	1.37e+01
32	73	29.2	379	8	Q47996	CYTOCHROME	1.37e+01
33	73	29.2	379	8	Q9XM21	CYTOCHROME	1.37e+01
34	73	29.2	379	8	Q9ZY15	CYTOCHROME	1.37e+01
35	73	29.2	379	8	Q48309	CYTOCHROME	1.37e+01
36	73	29.2	380	8	Q34P93	CYTOCHROME	1.37e+01
37	73	29.2	1409	5	Q026Z6	REGULATOR	1.37e+01
38	72	28.8	248	2	P74579	THIOL:DISU	1.80e+01
39	72	28.8	379	8	Q9XN83	CYTOCHROME	1.80e+01
40	72	28.8	380	8	Q79199	CYTOCHROME	1.80e+01
41	72	28.8	380	8	Q48063	CYTOCHROME	1.80e+01
42	72	28.8	414	1	Q96Z98	ASPARAGINY	1.80e+01
43	72	28.8	962	13	Q9W6N4	RECOMBINA	1.80e+01
44	72	28.8	1054	6	Q60577	EPHOE	1.80e+01
45	72	28.8	2873	14	Q94N1	POLYPR	1.80e+01

ALIGNMENTS

RESULT 1
ID O63620 PRELIMINARY; PRT; 645 AA.
AC O63620;
DT 01-AUG-1998 (TrEMBLrel. 02, Created)
DT 01-AUG-1998 (TrEMBLrel. 02, last sequence update)
DT 01-MAY 1999 (TrEMBLrel. 11, last annotation update)
DE NADH DEHYDROGENASE SUBUNIT 5.
OS Balanoglossus carnosus.
OG Mitochondrion.
OC Eukaryota; Metazoa; Bemich; Estata; Enteropneusta; -
OC Balanoglossus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98188267.
RA CASTRESANA J., FELDMAIER-FRITHS S., PAABO S.;
RT "Codon reassignment and amino acid composition in
RT mitochondria.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:3703-3707(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA CASTRESANA J., FELDMAIER-FRITHS S., PAABO S.;
RL Submitted (FEB 1998) to the EMBL/GenBank/DDBJ dat
DR EMBL; AF051097; AAD11945.1; -.
DR PFAM; PF00361; oxidored_q1; 1.
DR PFAM; PF00662; oxidored_q2; 1.
KW Mitochondrion.
SQ SEQUENCE 645 AA; [illegible] MW; [illegible] CRC64;

Query Match 37.6%; Score 94; DB 8; Length
Best Local Similarity 58.7%; Pred. No. 1.02e-02;
Matches 17; Conservative 5; Mismatches 6;

Db 45 PLIGAQLKF-INK-FLSIIHLTSPIISLV 72
P I :|| | |:| |||:|| : || |||
Qy 3 PAIPTQLYFPIDRSFLAILHQLILTTMSLV 32

RESULT 2
ID Q9Y6J9 PRELIMINARY; PRT; 622 AA.
AC Q9Y6J9;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, last annotation update)
DE PCAF ASSOCIATED FACTOR 65 ALPHA.

```
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98337182.
RA   OGRYZKO V.V., KOTANI T., ZHANG X., SCHILTZ R.L., HOWARD T.,
RA   YANG X.-J., HOWARD B.H., QIN J., NAKATANI Y., OGRYZKO V.V., KOTANI T.,
RA   ZHANG X., SCHILTZ R.L., HOWARD T., YANG X.-J., HOWARD B.H., QIN J.,
RA   NAKATANI Y.;
RT   "Histone-like TAFs within the PCAF histone acetylase complex.";
RL   Cell 94:35-44(1998).
DR   EMBL; AF069735; AAC39905.1; -.
SQ   SEQUENCE   622 AA;  67814 MW;  47EC7FA5 CRC32;

  Query Match           36.4%;  Score 91;  DB 4;  Length 622;
  Best Local Similarity  50.0%;  Pred. No. 3.03e-02;
  Matches   14;  Conservative    5;  Mismatches   7;  Indels   2;  Gaps   2;

Db      91 MRPAREGELYFPEDRE-VN-LVELALAT 116
           ||||   :|||| ||  :  | :|:|:|
Qy       1 MRPAIPTQLYFPTDRSFLAILHQLTLPT 28


RESULT    3
ID   Q9W6N6      PRELIMINARY;      PRT;   962 AA.
AC   Q9W6N6;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   RECOMBINATION ACTIVATING PROTEIN 1 (FRAGMENT).
GN   RAG-1.
OS   Gavialis gangeticus.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;
OC   Crocodylidae; Gavialinae; Gavialis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   GROTH J.G., BARROWCLOUGH G.F.;
RT   "Basal divergences in birds and the phylogenetic utility of the
RT   nuclear RAG-1 gene.";
RL   Mol. Phylogenet. Evol. 0:0-0(1999).
CC   -!- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
DR   EMBL; AF143725; AAD34951 1; -.
DR   PROSITE; PS00518; ZINC_FINGER_C3HC4, 1.
KW   Zinc-finger.
FT   NON_TER       1       1
FT   NON_TER     962     962
SQ   SEQUENCE   962 AA;  109642 MW;  713E78A7 CRC32;

  Query Match           32.8%;  Score 82;  DB 13;  Length 962;
  Best Local Similarity  58.3%;  Pred. No. 7.13e-01;
  Matches   14;  Conservative    2;  Mismatches   7;  Indels   1;  Gaps   1;

Db     303 PCFPTDLVTPV-RSFLNILDALTL 325
           | :|||:|  |  |||| ||  |||
Qy       3 PAIPTQLYFPTDRSFLAILHQLTL 26


RESULT    4
ID   Q23217      PRELIMINARY;      PRT;   434 AA.
AC   Q23217;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   W07G4.2 PROTEIN.
GN   W07G4.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   BAYNES C.;
RL   Submitted (AUG-1996) to the EMBL/GenBank/DDBJ databases.

RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE R., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z78018; CAB01447.1; -.
SQ   SEQUENCE   434 AA;  50304 MW;  A7BF2973 CRC32;

  Query Match           32.4%;  Score 81;  DB 5;  Length 434;
  Best Local Similarity  30.8%;  Pred. No. 1.00e+00;
  Matches    8;  Conservative   10;  Mismatches   8;  Indels   0;  Gaps   0;

Db     103 LFFVEDKTYLNMLNKVKEAIPGLIDQ 128
           |:|  |:::| :|: :  :  :|:||
Qy       9 LYFPIDRSFLAILHQLTLTTMSLVDQ 34


RESULT    5
ID   Q9YEJ5      PRELIMINARY;      PRT;   322 AA.
AC   Q9YEJ5;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   322AA LONG HYPOTHETICAL TRANSKETOLASE.
GN   APE0583.
OS   Aeropyrum pernix.
OC   Archaea; Crenarchaeota; Aeropyrum.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K1;
RX   MEDLINE; 99310339.
RA   KAWARABAYASI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y.,
RA   JIN-NO K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,
RA   HOSOYAMA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,
RA   TAKAMIYA M., MASUDA S., FUNAHASHI T., TANAKA T., KUDOH Y.,
RA   YAMAZAKI J., KUSHIDA N., OGURA A., ANKI K., KUBOTA K., NAKAMURA Y.,
RA   NOMURA N., SAKO Y., KIKUCHI H.;
RT   "Complete genome sequence of an aerobic hyper-thermophilic
RT   crenarchaeon, Aeropyrum pernix K1.";
RL   DNA Res. 6:83-101(1999).
DR   EMBL; AP000060; BAA79551.1; -.
SQ   SEQUENCE   322 AA;  34574 MW;  B42B3A33 CRC32;

  Query Match           31.6%;  Score 79;  DB 1;  Length 322;
  Best Local Similarity  24.2%;  Pred. No. 1.96e+00;
  Matches    8;  Conservative   13;  Mismatches  12;  Indels   0;  Gaps   0;

Db     282 RIGVPPGVYGSSSRDYMSLLRHLRLDPESVAAR 314
           | ::|  :| :: | ::::|::| |   |: :
Qy       2 PPAIPTQLYFPTDRSFLAILHQLTLTTMSLVDQ 34


RESULT    6
ID   P78998      PRELIMINARY;      PRT;   376 AA.
AC   P78998;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   HYPOTHETICAL 41.5 KD PROTEIN.
OS   Saccharomyces pastorianus (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RA   ANDERSEN T., NILSSON-TILLGREN T.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z86109; CAB06793.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   376 AA;  41542 MW;  BFF04E0D CRC32;

  Query Match           31.6%;  Score 79;  DB 3;  Length 376;
  Best Local Similarity  33.3%;  Pred. No. 1.96e+00;
  Matches   11;  Conservative   11;  Mismatches  10;  Indels   1;  Gaps   1;

Db    102 ISSSLPVQAYIDSDGRVAIPTLLDHWNPAIMGLI 134
          : :::| | |: :| |  |:|| :  :|:|:|
Qy      1 MRPAIPTQLYFPID-RSFLAILHQLILTTMSLV 32


RESULT    7
ID   Q9WTI1      PRELIMINARY;      PRT;   459 AA.
AC   Q9WTI1;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   ANTITHROMBIN III.
OS   Cavia porcellus (Guinea pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   SUZUKI Y.;
RT   "Guinea pig gene for antithrombin III.";
RL   Submitted (MAY 1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB027238; BAA77781.1; -.
SQ   SEQUENCE   459 AA;  51912 MW;  D7541DED CRC32;

  Query Match           31.2%;  Score 78;  DB 11;  Length 459;
  Best Local Similarity  27.3%;  Pred. No. 2.73e+00;
  Matches    9;  Conservative   14;  Mismatches  10;  Indels   0;  Gaps   0;

Db    420 RSLNPNRVIFKANRPPIVIIIFVTINTIIFMGP 452
          |:  | :: | ::|:|| :: :|:| |: :: :
Qy      2 RPAIPTQLYFPTDRSFLAILHQLILTTMSLVDQ 34


RESULT    8
ID   O59576      PRELIMINARY;      PRT;   144 AA.
AC   O59576;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   144AA LONG HYPOTHETICAL PROTEIN.
GN   PH1913.
OS   Pyrococcus horikoshii.
OC   Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OT3;
RX   MEDLINE; 98344137.
RA   KAWARABAYASI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
RA   YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOYAMA A., NAGAI Y.,
RA   SAKAI M., OGURA K., OTAKA A., NAKAMURA Y., SUZUKI H., DOI H.,
RA   FUNAHASHI T., TANAKA T., KUROKAWA Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,
RA   AOKI K., NAKAMURA Y., PODAR T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,
RA   KIKUCHI H.;
RT   "Complete sequence and gene organization of the genome of a hyper-
RT   thermophilic archaebacterium, Pyrococcus horikoshii OT3.";
RL   DNA Res. 5:55-76(1998).
DR   EMBL; AP000007; BAA31038.1; -.
SQ   SEQUENCE   144 AA;  15493 MW;  14C292B5 CRC32;

  Query Match           30.8%;  Score 77;  DB 1;  Length 144;
  Best Local Similarity  50.0%;  Pred. No. 3.79e+00;
  Matches    9;  Conservative    3;  Mismatches   6;  Indels   0;  Gaps   0;

Db    127 LRFTIPSKLYFTLENSEI 144
          :|  ||: ||| : :||
Qy      1 MRPAIPTQLYFPTDRSFL 18


RESULT    9
ID   O47997      PRELIMINARY;      PRT;   379 AA.
AC   O47997;
DT   01-JUN-1998 (TrEMBLrel. 05, Created)
DT   01-JUN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   CYTOCHROME B.
GN   CYTB.
OS   Thomomys umbrinus.
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;
OC   Eutheria; Rodentia; Sciuromorpha; Geomyidae; Thom
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MVZ 153878;
RX   MEDLINE; 98140315.
RA   SMITH M.F.;
RT   "Phylogenetic relationships and geographic struct
RT   in the genus Thomomys.";
RL   Mol. Phylogenet. Evol. 9:1-11(1998).
CC   -!- CATALYTIC ACTIVITY: QH(2) + 2 FERRICYTOCHROME
CC       FERROCYTOCHROME C.
CC   -!- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHI
CC       BOUND TO THE PROTEIN (BY SIMILARITY).
DR   EMBL; U65287; AAC40025.1; -.
DR   PFAM; PF00032; cytochrome_b_C; 1.
DR   PFAM; PF00033; cytochrome_b_N; 1.
KW   Mitochondrion; Electron transport; Respiratory cha
KW   Heme.
SQ   SEQUENCE   379 AA;  43115 MW;  4E98FAA3 CRC32;

  Query Match           30.8%;  Score 77;  DB 8;  Leng
  Best Local Similarity  42.9%;  Pred. No. 3.79e+00;
  Matches   12;  Conservative    5;  Mismatches   9;

Db    218 IPFHXYYXI-KDFLGAI-ILIIFFMIV 243
          || : |: | : || |  |  :: |
Qy      5 IPTQLYFPTDRSFLAILHQL:::MSLV 32


RESULT   10
ID   Q36229      PRELIMINARY;      PRT;   379 AA.
AC   Q36229;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update
DE   CYTOCHROME B.
GN   CYT B.
OS   Varecia variegata rubra.
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;
OC   Eutheria; Primates; Strepsirhini; Lemuridae; Vare
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96229867.
RA   YODER A.D., CARTMILL M., RUVOLO M., SMITH K., VILG
RT   "Ancient single origin for Malagasy primates.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:5122-5126(1996).
CC   -!- CATALYTIC ACTIVITY: QH(2) + 2 FERRICYTOCHROME
CC       FERROCYTOCHROME C.
CC   -!- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHI
CC       BOUND TO THE PROTEIN (BY SIMILARITY).
DR   EMBL; U53578; AAC50544.1; -.
DR   PFAM; PF00032; cytochrome_b_C; 1.
DR   PFAM; PF00033; cytochrome_b_N; 1.
KW   Mitochondrion; Electron transport; Respiratory cha
KW   Heme.
```

```
SQ   SEQUENCE   479 AA;  43003 MW;  61DF7365 CRC32;

  Query Match             30.4%;  Score 76;  DB 8;  Length 479;
  Best Local Similarity   42.9%;  Pred. No. 5.25e+00;
  Matches   12;  Conservative    7;  Mismatches   7;  Indels   2;  Gaps   2;

Db    218 IPFHPYY-TIKDFLGLL-LLILLTVILV 243
          || : |: | : |||:|  | | |:||
Qy      5 IPTQLYFPTDRSFLAILHQLTLTMSLV 32


RESULT   11
ID   O18233      PRELIMINARY;      PRT;   571 AA
AC   O18233;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   Y57G11C.7 PROTEIN.
GN   Y57G11C.7.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   MCMURRAY A.;
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z99281; CAB16509.1; -.
SQ   SEQUENCE   571 AA;  66884 MW;  2A34A731 CRC32;

  Query Match             30.4%;  Score 76;  DB 5;  Length 571;
  Best Local Similarity   35.7%;  Pred. No. 5.25e+00;
  Matches   10;  Conservative   10;  Mismatches   6;  Indels   2;  Gaps   2;

Db    259 QINPQSEKSPLSHPLQNPLSYLKTTNPNE 286
          || | :::|||: :|:  |:| | :: :
Qy      8 QLYFPTDRSFLA-ILHQ LTLTTMSLVD 33


RESULT   12
ID   Q9W6N5      PRELIMINARY;      PRT;   962 AA.
AC   Q9W6N5;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   RECOMBINATION ACTIVATING PROTEIN 1 (FRAGMENT).
GN   RAG-1.
OS   Alligator mississippiensis (American alligator).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;
OC   Crocodylidae; Alligatorinae; Alligator.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   GROTH J.G., BARROWCLOUGH G.F.;
RT   "Basal divergences in birds and the phylogenetic utility of the
RT   nuclear RAG-1 gene.";
RL   Mol. Phylogenet. Evol. 0:0-0(1999).
CC   -!- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
DR   EMBL; AF143724; AAD34950.1; -.
DR   PROSITE; PS00518; ZINC_FINGER_C3HC4; 1.
KW   Zinc finger.
FT   NON_TER       1      1
FT   NON_TER     962    962
SQ   SEQUENCE   962 AA;  109816 MW;  DA3A4902 CRC32;

  Query Match             30.4%;  Score 76;  DB 13;  Length 962;
  Best Local Similarity   54.2%;  Pred. No. 5.25e+00;
  Matches   13;  Conservative    3;  Mismatches   7;  Indels   1;  Gaps   1;

Db    303 PCFPTDDLVTPV-KSFLNILDSLTL 325
          | :||:|  |  :||| ||  |||
Qy      3 PAIPTQLYFPTDRSFLAILHQLTL 26


RESULT   13
ID   Q94620      PRELIMINARY;      PRT;  1209 AA.
AC   Q94620;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   F27D4.6 PROTEIN.
GN   F27D4.6.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   WILKINSON J.;
RL   Submitted (SEP-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z79695; CAB01971.1; -.
SQ   SEQUENCE   1209 AA;  139656 MW;  F701D116 CRC32;

  Query Match             30.4%;  Score 76;  DB 5;  Length 1209;
  Best Local Similarity   47.6%;  Pred. No. 5.25e+00;
  Matches   10;  Conservative    5;  Mismatches   6;  Indels   0;  Gaps   0;

Db    541 FTTIQIFIPVLQKLTLQTFSL 561
          | | : |:::|: |||  | ||
Qy     11 FPTDRSFLAILHQLTLTTMSL 31


RESULT   14
ID   O41977      PRELIMINARY;      PRT;  1275 AA.
AC   O41977;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   TEGUMENT PROTEIN/FGARAT.
GN   GAMMAHV.ORF75B.
OS   murine herpesvirus 68.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WUMS;
RX   MEDLINE; 97366649.
```

```
RA   VIRGIN H.W. IV, LATREILLE P., WAMSLEY P., HALLSWORTH K., WECK K.E.,
RA   DAL CANTO A.J., SPECK S.H.;
RT   "Complete sequence and genomic analysis of murine gammaherpesvirus
RT   68.";
RL   J. Virol. 71:5894-5904(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WUMS;
RA   LATREILLE P., WAMSLEY P., WATERSTON R.H.;
RL   Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U97553; AAB66459.1; -.
SQ   SEQUENCE   1275 AA;  141953 MW;  518C97E0 CRC32;

  Query Match           30.4%;  Score 76;  DB 14;  Length 1275;
  Best Local Similarity  36.0%;  Pred. No. 5.25e+00;
  Matches    9;  Conservative    8;  Mismatches   8;  Indels   0;  Gaps   0;

Db    310 IRPISSNQLIFPIFVNFVNFISKLS 334
          :||  : || ||| :|: ::  |:
Qy      1 MRPAIPTQLYFPTDRSFLAILHQLI 25


RESULT   15
ID   O74390       PRELIMINARY;      PRT;  2609 AA.
AC   O74390;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV 1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV 1999 (TrEMBLrel. 12, Last annotation update)
DE   PUTATIVE TRANSPORT PROTEIN.
GN   SPBC3H7.16 OR SPBC28E12.06C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Archiascomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972H-;
RA   LYNE M., RAJANDREAM M.A., BARRELL B.G., JIMENEZ MARTINEZ J;
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 2303-2609 FROM N.A.
RC   STRAIN=972H-;
RA   WOOD V., RAJANDREAM M.A., BARRELL B.G., SKELTON J., CHURCHER C.M.;
RL   Submitted (SEP 1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL031261; CAA20312.1; -.
DR   EMBL; AL031517; CAA20654.1; -.
DR   PFAM; PF00400; WD40; 2.
DR   PRINTS; PR00320; GPROTEINBRPT.
SQ   SEQUENCE   2609 AA;  295654 MW;  BB204170 CRC32;

  Query Match           30.4%;  Score 76;  DB 3;  Length 2609;
  Best Local Similarity  42.9%;  Pred. No. 5.25e+00;
  Matches    9;  Conservative    5;  Mismatches   6;  Indels   1;  Gaps   1;

Db   1837 LR-DVGLELFFKDGRSFLFIL 1856
          :|  :  :|:|   |||:||
Qy      1 MRPAIPTQLYFPTDRSFLAIL 21


Search completed: Fri Aug 25 18:06:13 2000
Job time : 52 secs.
```

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 25 18:01:35 2000; MasPar time 7.31 Seconds
332.083 Million cell updates/sec

Tabular output not generated.

Title: >US-09-257-585-6
Description: (1-35) from US09257585.pep
Perfect Score: 244
Sequence: 1 MDASLNTILQQVDLIGYLLVISSILFGHYIALLAA 35

Scoring table: PAM 150
Gap 11

Searched: 225878 seqs, 69344122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl12
1:sp_archea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 34.187; Variance 68.711; scale 0.498

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	92	37.7	521	10	Q41301	BETA KETOACYL COA SYNT	8.09e-02
2	87	35.7	305	2	O67090	PHOSPHOTIDYLGLYCEROPHO	4.09e-01
3	86	35.2	268	2	O86889	CAP33FF PROTEIN.	5.63e-01
4	86	35.2	269	5	Q77389	PPC07250 PROTEIN.	5.63e-01
5	85	34.8	185	8	Q35531	CYTOCHROME OXIDASE SUB	7.72e-01
6	85	34.8	350	10	O81850	PUTATIVE FE(II) TRANSP	7.72e-01
7	85	34.8	418	2	Q56913	TRSA.	7.72e-01
8	84	34.0	581	11	O35120	MGP11P.	1.45e+00
9	84	34.0	581	11	O35456	N-ACETYLGLUCOSAMYL TRA	1.45e+00
10	84	34.0	597	5	Q27170	K+ CHANNEL.	1.45e+00
11	82	33.6	98	8	Q9X110	NADH DEHYDROGENASE SUB	1.97e+00
12	82	33.6	245	2	Q52260	PLASMID PED208 SURFACE	1.97e+00
13	81	33.2	386	10	O64737	HOOKLESS1-LIKE PROTEIN	2.68e+00
14	81	33.2	469	1	O30094	CONSERVED HYPOTHETICAL	2.68e+00
15	81	33.2	503	2	Q45934	STARCH DEGRADING ENZYM	2.68e+00
16	81	33.2	511	2	P94618	PUTATIVE ALPHA-AMYLASE	2.68e+00
17	79	32.4	504	4	Q99624	TRANSPORTER PROTEIN.	4.94e+00
18	79	32.4	684	2	Q51415	ALGINATE LYASE.	4.94e+00
19	78	32.0	74	2	Q56009	NADH DEHYDROGENASE SUB	6.68e+00
20	78	32.0	174	8	Q95961	CYTOCHROME OXIDASE SUB	6.68e+00
21	78	32.0	174	8	[illegible]	CYTOCHROME	[illegible]
22	78	32.0	180	8	[illegible]	CYTOCHROME	[illegible]
23	78	32.0	185	8	Q36080	CYTOCHROME	[illegible]
24	78	32.0	239	2	[illegible]	UNIDENTIFIE	[illegible]
25	78	32.0	581	1	[illegible]	N ACETYLGLU	[illegible]
26	78	32.0	709	5	[illegible]	T02B5.1 PRO	[illegible]
27	78	32.0	826	5	Q21664	SIMILAR TO	[illegible]
28	77	31.6	181	5	[illegible]	ADP-RIB-SYL	[illegible]
29	76	31.1	107	1	[illegible]	ORF C25G2.	[illegible]
30	76	31.1	220	2	[illegible]	NORD.	[illegible]
31	76	31.1	229	2	Q41882	CALLED ORF	[illegible]
32	76	31.1	621	2	Q48408	CELLOBIOSE-	[illegible]
33	76	31.1	864	2	P70811	ORF1.	[illegible]
34	75	30.7	174	8	Q95978	CYTOCHROME	[illegible]
35	75	30.7	179	2	[illegible]	PURINE BIND	[illegible]
36	75	30.7	185	8	[illegible]	CYTOCHROME	[illegible]
37	75	30.7	510	2	[illegible]	CPSI PROTEI	[illegible]
38	75	30.7	445	11	[illegible]	BLUE CONE	[illegible]
39	75	30.7	416	9	Q92X31	ORF22.	[illegible]
40	75	30.7	574	5	[illegible]	T02G6.5 PRO	[illegible]
41	75	30.7	1846	5	[illegible]	INSULIN REC	[illegible]
42	74	30.3	98	8	[illegible]	NADH DEHYDR	[illegible]
43	74	30.3	108	2	[illegible]	THREONINE S	[illegible]
44	74	30.3	448	8	[illegible]	BLUE OPSIN	[illegible]
45	74	30.3	458	10	[illegible]	PUTATIVE PE	[illegible]

ALIGNMENTS

RESULT 1
ID Q41301 PRELIMINARY; PRT; 521 AA.
AC Q41301;
DT 01-NOV-1996 (TrEMBLrel. 03, Created)
DT 01-NOV-1996 (TrEMBLrel. 03, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE BETA-KETOACYL-COA SYNTHASE.
OS Simmondsia chinensis (Jojoba).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudi
OC core eudicots; Caryophyllidae; Caryophyllales; Sim
OC Simmondsia.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96451420.
RA LASSNER M., LARDIZABAL K., METZ J.G.;
RT "A jojoba beta-Ketoacyl-CoA synthase cDNA complem
RT acid elongation mutation in transgenic plants.";
RL Plant Cell 8:281-292(1996).
DR EMBL; U37088; AAC49186.1;
DR MENDEL; 15119; Simch;2005;15119.
SQ SEQUENCE 521 AA; 58569 MW; 8045BC08 CRC32;

Query Match 37.7%; Score 92; DB 10; Len
Best Local Similarity 44.8%; Pred. No. 8.09e-02;
Matches 13; Conservative 9; Mismatches 7; G

Db 32 SINLHHVKL GYHYLISNALKVEIPLLG 59
Qy 6 NILQQVDLIGYLLVISSILFGHYIALLA 34

RESULT 2
ID O67090 PRELIMINARY; PRT; 305 AA.
AC O67090;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE PHOSPHOTIDYLGLYCEROPHOSPHATE SYNTHASE.
GN PGSA2.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquif. aceae; Aquifex.
RN [1]
RP SEQUENCE FROM N.A.

```
RC   STRAIN=VF5;
RX   MEDLINE; 98196666.
RA   DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
RA   GRAHAM D.E., OVERBEEK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
RA   FEILDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RT   "The complete genome of the hyperthermophilic bacterium Aquifex
RT   aeolicus.";
RL   Nature 392:353-358(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VF5;
RA   DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
RA   GRAHAM D.E., OVERBEEK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
RA   FEILDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RL   Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE000715; AAC07048.1; -.
DR   PROSITE; PS00379; CDP_ALCOHOL_P_TRANSF; 1.
DR   PFAM; PF01066; CDP-OH_P_transf; 1.
SQ   SEQUENCE   305 AA;  35265 MW;  4004CF793 CRC32;

  Query Match           35.7%;  Score 87;  DB 2;  Length 305;
  Best Local Similarity  34.5%;  Pred. No. 4.09e-01;
  Matches   10;  Conservative   12;  Mismatches   5;  Indels   2;  Gaps   2;

Db    125 SAISLILLGVELGNYLIIIA-VIFA-YIS 151
          ::::::|  |:|  |||:::  :|: ||:
Qy      3 ASLNIILQQVDLTGYLLVISSTLFGHYIA 31


RESULT    3
ID   O86889      PRELIMINARY;      PRT;   268 AA.
AC   O86889;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   CAP33FF PROTEIN.
GN   CAP33FF.
OS   Streptococcus pneumoniae.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SSISP 33F TYPE 33F;
RA   LLULL D., LOPEZ R., GARCIA E., MUNOZ R.;
RT   "Nucleotide sequence of the capsular gene cluster encoding the
RT   polysaccharide capsule of Streptococcus pneumoniae type 33F.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ006986; CAA07399.1; -.
SQ   SEQUENCE   268 AA;  31754 MW;  A0A41470 CRC32;

  Query Match           35.2%;  Score 86;  DB 2;  Length 268;
  Best Local Similarity  44.9%;  Pred. No. 5.63e-01;
  Matches   13;  Conservative    6;  Mismatches   9;  Indels   1;  Gaps   1;

Db    195 LDTMYEQMDLTDYTL-FESRLFGRVSELL 222
          |: : :|:||| | | : | |||:   ||
Qy      5 LNIILQQVDLTGYLLVISSTLFGHYIALL 33


RESULT    4
ID   O77389      PRELIMINARY;      PRT;   269 AA.
AC   O77389;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   PFC0725C PROTEIN.
GN   PFC0725C.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RA   MURPHY L., LAWSON D., BARRELL B.;
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z98551; CAB11145.1; -.
DR   PFAM; PF01226; Form_Nir_trans; 1.
SQ   SEQUENCE   269 AA;  29658 MW;  541A89A0 CRC32;

  Query Match           35.2%;  Score 86;  DB 5;  Length 269;
  Best Local Similarity  29.0%;  Pred. No. 5.63e-01;
  Matches    9;  Conservative   10;  Mismatches  12;  Indels   0;  Gaps   0;

Db    110 LAVTMALYEKKVKLLDYLRVMTISLFGNYVG 140
          :  :: :    | |  || |:: :|||:|::
Qy      1 MDASLNIILQQVDLTGYLLVISSTLFGHYIA 31


RESULT    5
ID   Q35531      PRELIMINARY;      PRT;   185 AA.
AC   Q35531;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   CYTOCHROME OXIDASE SUBUNIT II (FRAGMENT).
OS   Prodoxus marginatus.
OC   Mitochondrion.
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Monotrysia;
OC   Incurvarioidea; Prodoxidae; Prodoxus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SANTA BARBARA CO., CA;
RA   BROWN J.M., PELLMYR O., THOMPSON J.N., HARRISON R.G.;
RL   Submitted (JAN-1994) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U04886; AAA16080.1; -.
DR   PFAM; PF00116; COX2; 1.
KW   Mitochondrion.
FT   NON_TER     185     185
SQ   SEQUENCE   185 AA;  21614 MW;  0D3F5385 CRC32;

  Query Match           34.8%;  Score 85;  DB 8;  Length 185;
  Best Local Similarity  44.8%;  Pred. No. 7.72e-01;
  Matches   13;  Conservative    5;  Mismatches   9;  Indels   2;  Gaps   2;

Db     25 DHTLMILLMITILVGYLL-IT-ILFNNYV 51
          | :| |:|    | |||| |:  || :|:
Qy      2 DASLNIILQQVDLTGYLLVISSTLFGHYI 30


RESULT    6
ID   O81850      PRELIMINARY;      PRT;   350 AA.
AC   O81850;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   PUTATIVE FE(II) TRANSPORT PROTEIN.
GN   T16H5.40.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC   Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   DE HAAN M., MAARSE A.C., GRIVELL L.A., BANCROFT I., MEWES H.W.,
RA   MAYER K., SCHUELLER C., BEVAN M.;
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU ARABIDOPSIS SEQUENCING PROJECT;
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL024486; CAA19685.1; -.
DR   MENDEL; 32480; Arath;1756;32480.
SQ   SEQUENCE   350 AA;  37407 MW;  890F2506 CRC32;

  Query Match           34.8%;  Score 85;  DB 10;  Length 350;
```

```
  Best Local Similarity  36.7%;  Pred. No. 7.72e-01;
  Matches   11;  Conservative   10;  Mismatches   9;  Indels   0;  Gaps   0;

Db      45 PLKIVAIVAILLISLIGVISPLFSKYISFL 74
           :|:||:      || |: ::| ||:::|||:|
Qy       4 SLNIILQQVDLIGYLLVISSILFGHYIALL 33


RESULT    7
ID   Q56913       PRELIMINARY;      PRT;   418 AA
AC   Q56913;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   TRSA.
GN   TRSA.
OS   Yersinia enterocolitica.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Yersinia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN 6471/76 (SEROTYPE O:3);
RX   MEDLINE; 96100456.
RA   SKURNIK M., VENHO R., TOIVANEN P., AL HENDY A.;
RT   "A novel locus of Yersinia enterocolitica serotype O:3 involved in
RT   lipopolysaccharide outer core biosynthesis.";
RL   Mol. Microbiol. 17:575-594(1995).
DR   EMBL; X47767; CAA87698.1; -.
SQ   SEQUENCE   418 AA;  47170 MW;  7300265B CRC32;

  Query Match           34.8%;  Score 85;  DB 2;  Length 418;
  Best Local Similarity  66.7%;  Pred. No. 7.72e-01;
  Matches   12;  Conservative    4;  Mismatches   1;  Indels   1;  Gaps   1;

Db     333 YLLVIQSTVLGNYI-LLP 349
           ||||| ||::|:|| ||
Qy      17 YLLVISSTLFGHYIALLA 34


RESULT    8
ID   O35120       PRELIMINARY;      PRT;   581 AA.
AC   O35120;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JAN-1998 (TrEMBLrel. 05, Last annotation update)
DE   MGPIIP.
GN   MGPII.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RA   HONG Y.;
RT   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB008895; BAA23615.1; -.
SQ   SEQUENCE   581 AA;  66204 MW;  153CDEF3 CRC32;

  Query Match           34.0%;  Score 83;  DB 11;  Length 581;
  Best Local Similarity  34.3%;  Pred. No. 1.45e+00;
  Matches   12;  Conservative   12;  Mismatches  10;  Indels   1,  Gaps   1,

Db     362 MSPFIEHILWHVGLSACLGLTVALSIFSDIIALLT 396
           | : :: || :| |:: | | :: ::|:  ||||:
Qy       1 MDASLNIILQQVDLIGYL-LVISSILFGHYIALLA 34


RESULT    9
ID   O35456       PRELIMINARY;      PRT;   581 AA.
AC   O35456;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   N ACETYLGLUCOSAMYL TRANSFERASE COMPONENT GPI1.
GN   GPI1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;
OC   Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=BRAIN;
RA   TIEDE A., SCHUBERT J., ORLEAN P., SCHMIDT R.E.;
RT   Submitted (OCT-1997) to the EMBL/GenBank/DDBJ data
DR   EMBL; AF030178; AAC79674.1; -.
KW   Transferase.
SQ   SEQUENCE   581 AA;  65909 MW;  40CB968C CRC32;

  Query Match           34.0%;  Score 83;  DB 11;  Le
  Best Local Similarity  34.3%;  Pred. No. 1.45e+00;
  Matches   12;  Conservative   12;  Mismatches  10;

Db     362 MSPFIEHILWHVGLSACLGLTVALSIFSDIIALLT 396
           | : :: || :| |:: | | :: ::|:  ||||:
Qy       1 MDASLNIILQQVDLIGYL-LVISSILFGHYIALLA 34


RESULT   10
ID   Q27170       PRELIMINARY;      PRT;   597 AA.
AC   Q27170;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   K+ CHANNEL.
GN   PAK2.
OS   Paramecium tetraurelia.
OC   Eukaryota; Alveolata; Ciliophora; Nassophorea; Pe
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96172462.
RA   JEGLA T., SALKOFF L.;
RT   "A multigene family of novel K+ channels from Para
RT   tetraurelia.";
RL   Recept. Channels 3:51-63(1995).
DR   EMBL; U19908; AAA95997.1;
DR   PFAM; PF00914; CNG_membrane; 1.
KW   Ionic channel.
SQ   SEQUENCE   597 AA;  [illegible] MW;  B9F2E186 CRC32;

  Query Match           34.0%;  Score 83;  DB 5;  Len
  Best Local Similarity  42.9%;  Pred. No. 1.45e+00;
  Matches   12;  Conservative    7;  Mismatches   9;

Db     495 LINILQQNDPSNYLYILSNGSLSFYIIL 522
           |  |||| | : || :::  :: |:::
Qy       5 LNIILQQVDLIGYLLVISSILFGHYIAL 32


RESULT   11
ID   Q9XL10       PRELIMINARY;      PRT;    98 AA.
AC   Q9XL10;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update
DE   NADH DEHYDROGENASE SUBUNIT IV I.
OS   Falco peregrinus.
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;
OC   Neognathae; Falconiformes; Falconidae; Falco.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98393707.
RA   MINDELL D.P., SORENSON M.D., HUELSENBECK C.J., MIRA
RA   KNIGHT A., SAWCHUK S.J., YURI T.;
RT   "Multiple independent origins of mitochondrial gen
RL   (in) Mindell D.P. (eds.);
RL   Avian molecular evolution and systematics, pp.211
```

```
RL   New York (1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98393707.
RA   MINDELL D.P., SORENSON M.D., DIMCHEFF D.E.;
RT   "Multiple independent origins of mitochondrial gene order in birds.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:10693-10697(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   MINDELL D.P., SORENSON M.D., DIMCHEFF D.E., HASEGAWA M., AST J.C.,
RA   YURI T.;
RT   "Interordinal relationships of birds and other reptiles based on whole
RT   mitochondrial genomes.";
RL   Syst. Biol. 48:138-152(1999).
RN   [4]
RP   SEQUENCE FROM N.A.
RA   SORENSON M.D., DIMCHEFF D.E., AST J.C., YURI T., MINDELL D.P.;
RT   "Complete mitochondrial DNA sequences for five birds and a turtle.";
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF090338; AAD32500.1; -.
KW   Mitochondrion.
SQ   SEQUENCE   98 AA;  10478 MW;  C37A9271 CRC32;

  Query Match           33.6%;  Score 82;  DB 8;  Length 98;
  Best Local Similarity  36.7%;  Pred. No. 1.97e+00;
  Matches   11;  Conservative    9;  Mismatches  10;  Indels   0;  Gaps   0;

Db      15 SSLGLTLHRTHLISALLCLESMMLSMYIAL 44
           :|| : |::  | : || : | ::: ||||
Qy       3 ASLNTILQQVDLTGYLLVISSTLFGHYIAL 32


RESULT   12
ID   Q52260      PRELIMINARY;      PRT;   245 AA.
AC   Q52260;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   PLASMID PED208 SURFACE EXCLUSION SYSTEM
DE   (SFX) GENES TRAS AND TRAT, ENCODING TRATP (FRAGMENT).
GN   TRAD.
OS   unidentified bacterium.
OG   Plasmid IncFV pED208.
OC   Bacteria; environmental samples.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 86223783.
RA   FINLAY B.B., PARANCHYCH W.;
RT   "Nucleotide sequence of the surface exclusion genes traS and traT from
RT   the IncFO lac plasmid pED208.";
RL   J. Bacteriol. 166:713-721(1986).
DR   EMBL; M13465; AAA88376.1; -.
KW   Plasmid.
FT   NON_TER     245     245
SQ   SEQUENCE   245 AA;  28363 MW;  C9A85EDC CRC32;

  Query Match           32.6%;  Score 82;  DB 2;  Length 245;
  Best Local Similarity  31.0%;  Pred. No. 1.97e+00;
  Matches    9;  Conservative   12;  Mismatches   8;  Indels   0;  Gaps   0;

Db       3 LKMFLQINNLISYYVIMGTVLFGIAVLLM 31
           |:::||  :| :| ::::: ||| : |:
Qy       5 LNTILQQVDLTGYLLVISSTLFGHYIALL 33


RESULT   13
ID   O64737      PRELIMINARY;      PRT;   386 AA.
AC   O64737;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   HOOKLESS1-LIKE PROTEIN.
GN   T27E13.17.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta,
OC   euphyllophytes, Spermatophyta, Magnoliophyta, eudicotyledons,
OC   core eudicots, Rosidae, eurosids II, Brassicales, Brassicaceae;
OC   Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   BOUNSLEY S.D., LIN X., KETCHUM K.A., CROSBY M.I., BRANDON R.C.,
RA   SYKES S.M., KAUL S., MASON T.M., FERLAVAGE A.P., ADAMS M.D.,
RA   SOMERVILLE C.R., VENTER J.C.;
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC004165; AAC16963.1; -.
DR   MENDEL; 28660; Arath;1740;28660.
DR   PFAM; PF00583; Acetyltransf; 1.
SQ   SEQUENCE   386 AA;  43917 MW;  7848C776 CRC32;

  Query Match           33.2%;  Score 81;  DB 10;  Length 386;
  Best Local Similarity  32.3%;  Pred. No. 2.68e+00;
  Matches   10;  Conservative   13;  Mismatches   6,  Indels   2,  Gaps   2;

Db     243 FKLRIERAPLS-YLLLTKVSKLFGNFLSLLG 272
           ::: ::: |: |||:  | |||:::||:
Qy       5 LNTILQQVDLTGYLLVIS-STLFGHYIALLA 34


RESULT   14
ID   O30094      PRELIMINARY;      PRT;   469 AA.
AC   O30094;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   CONSERVED HYPOTHETICAL PROTEIN.
GN   AF0143.
OS   Archaeoglobus fulgidus.
OC   Archaea, Euryarchaeota, Archaeoglobales, Archaeoglobaceae;
OC   Archaeoglobus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN VC-16 / DSM 4304 / ATCC 49558;
RX   MEDLINE; 98049343.
RA   KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
RA   KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,
RA   RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRPIDES N.C.,
RA   FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
RA   KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
RA   PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L.,
RA   OVERBEEK R., GOCAYNE J.D., WEIDMAN J.F., MCDONALD L., UTTERBACK T.,
RA   COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M.,
RA   SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
RA   MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,
RA   VENTER J.C.;
RT   "The complete genome sequence of the hyperthermophilic, sulphate-
RT   reducing archaeon Archaeoglobus fulgidus.";
RL   Nature 390:364-370(1997).
DR   EMBL; AE001096; AAB91081.1; -.
DR   1IGK, AF0143, -.
KW   Hypothetical protein.
SQ   SEQUENCE   469 AA;  52627 MW;  E0853A88 CRC32;

  Query Match           33.2%;  Score 81;  DB 1;  Length 469;
  Best Local Similarity  38.1%;  Pred. No. 2.68e+00;
  Matches    8;  Conservative    8;  Mismatches   5;  Indels   0;  Gaps   0;

Db      31 FLLTAINLSVFLLIIFTGLFG 51
           ::|  ::|: :|||: | : |||
Qy       7 LTLQQVDLTGYLLVISSTLFG 27

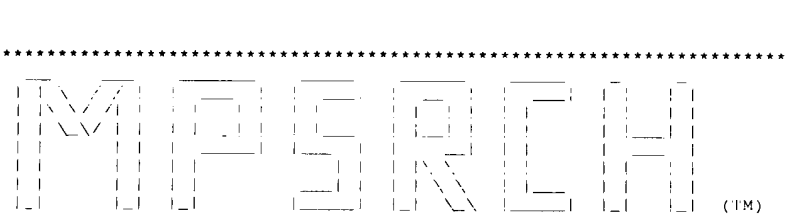

RESULT   15
ID   Q45834      PRELIMINARY;      PRT;   503 AA.
AC   Q45834;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
```

```
DT   01-NOV-1996 (IrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (IrEMBLrel. 08, Last annotation update)
DE   STARCH DEGRADING ENZYME PRECURSOR.
OS   Clostridium acetobutylicum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 824;
RA   VERHASSELT P., VANDERLEYDEN J., VAN LEUVEN F.;
RL   Submitted (FEB-1993) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X70834; CAA49800.1; -.
DR   PFAM; PF00877; NLPC_P60; 1.
KW   Signal.
FT   SIGNAL        1     28       POTENTIAL.
FT   CHAIN        29    503       STARCH DEGRADING.
SQ   SEQUENCE   503 AA;  53930 MW;  476DA117 CRC32;

  Query Match             33.2%;  Score 81;  DB 2;  Length 503;
  Best Local Similarity   38.5%;  Pred. No. 2.68e+00;
  Matches    10;  Conservative    8;  Mismatches   8;  Indels   0;  Gaps   0;

Db       1 MKKIFTNVLLVGFLVFISAILYGHSV 26
           :: |:  | | |:|: ||: |:|| :
Qy       5 LNIILQQVDLTGYLLVISSTLFGHYI 30


Search completed: Fri Aug 25 18:02:37 2000
Job time : 62 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 25 18:01:06 2000; MasPar time 3.40 Seconds
418.631 Million cell updates/sec

Tabular output not generated.

Title: >US-09-257-585-6
Description: (1-35) from US09257585.pep
Perfect Score: 244
Sequence: 1 MDASLNIILQQVDLIGYLLVISSTLFGHYIALLAA 35

Scoring table: PAM 150
Gap 11

Searched: 85661 seqs, 30989116 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 35.628; Variance 64.140; scale 0.555

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	87	35.7	361	1	SERC_BUCAP	PHOSPHOSERINE AMINOTRA	9.59e-02
2	83	34.0	165	1	MLCH_RAT	MELANIN-CONCENTRATING	3.76e-01
3	82	33.6	165	1	MLCH_HUMAN	MELANIN-CONCENTRATING	5.25e-01
4	79	32.4	71	1	MLCH_HYLLA	MELANIN CONCENTRATING	1.41e+00
5	79	32.4	71	1	MLCH_PANTR	MELANIN-CONCENTRATING	1.41e+00
6	79	32.4	71	1	MLCH_PANPA	MELANIN CONCENTRATING	1.41e+00
7	79	32.4	186	1	SLFA_PSEAE	NADH-DEPENDENT FMN RED	1.41e+00
8	79	32.4	473	1	PSBC_CYACA	PHOTOSYSTEM II 44 KDA	1.41e+00
9	78	32.0	343	1	Y841_METTH	HYPOTHETICAL PROTEIN M	1.96e+00
10	77	31.6	273	1	TRPA_PORPU	TRYPTOPHAN SYNTHASE AL	2.70e+00
11	77	31.6	346	1	OPSB_MOUSE	BLUE-SENSITIVE OPSIN (	2.70e+00
12	77	31.6	346	1	OPSB_RAT	BLUE-SENSITIVE OPSIN (	2.70e+00
13	76	31.1	139	1	FXSA_SERMA	FXSA PROTEIN (FRAGMENT	3.71e+00
14	76	31.1	262	1	PCYA_ACRPB	PROTEIN-TYROSINE PHOSP	3.71e+00
15	76	31.1	292	1	CITG_ECOLI	CITG PROTEIN.	3.71e+00
16	76	31.1	527	1	NU5M_CAEEL	NADH-UBIQUINONE OXIDOR	3.71e+00
17	75	30.7	486	1	MURD_MYCTU	UDP-N-ACETYLMURAMOYLAL	5.09e+00
18	75	30.7	488	1	FA10_HUMAN	COAGULATION FACTOR X P	5.09e+00
19	75	30.7	522	1	NU5M_MYTED	NADH-UBIQUINONE OXIDOR	5.09e+00
20	74	30.3	203	1	YH12_CAMJE	HYPOTHETICAL 23.2 KDA	6.96e+00
21	74	30.3	348	1	OPSB_HUMAN	BLUE-SENSITIVE OPSIN (	6.96e+00
22	74	30.3	473	1	PSBC_SECCE	PHOTOSYSTEM II 44 KDA	6.96e+00
23	74	30.3	473	1	PSBC_HORVU	PHOTOSYSTEM II 44 KDA	6.96e+00
24	73	29.9	71	1	MLCH_TARSY	MELANIN-CON	[illegible]
25	73	29.9	140	1	Y314_YEAST	HYPOTHETICA	[illegible]
26	73	29.9	295	1	MOTA_SALTY	CHEMOTAXIS P	[illegible]
27	73	29.9	295	1	YMB6_YEAST	HYPOTHETICA	[illegible]
28	73	29.9	473	1	PSBC_TOBAC	PHOTOSYSTEM	[illegible]
29	72	29.5	296	1	BIPX_MYCTU	PROBABLE PR	[illegible]
30	72	29.5	344	1	MELA_SHECO	MELA PROTEIN	[illegible]
31	72	29.5	473	1	PSBC_MAIZE	PHOTOSYSTEM	[illegible]
32	72	29.5	564	1	ESTJ_HELVI	JUVENILE HO	[illegible]
33	72	29.5	834	1	MPE5_YEAST	SUPPRESSOR	[illegible]
34	72	29.5	919	1	YKU5_YEAST	HYPOTHETICA	[illegible]
35	72	29.5	1078	1	CRM1_SCHPO	CHROMOSOME	[illegible]
36	71	29.1	173	1	YO96_NPVAC	HYPOTHETICA	[illegible]
37	71	29.1	214	1	ATP6_ALBCO	ATP SYNTHAS	[illegible]
38	71	29.1	297	1	PEPM_ERWCH	PEPM PROTEI	[illegible]
39	71	29.1	412	1	SGAT_HAEIN	SERINE TRAN	[illegible]
40	71	29.1	433	1	RODA_TREPA	ROD SHAPE-D	[illegible]
41	71	29.1	459	1	RADA_THLPN	DNA REPAIR	[illegible]
42	71	29.1	487	1	PSBC_PINPU	PHOTOSYSTEM	[illegible]
43	71	29.1	687	1	CYSA_HELPY	CARBON STAR	[illegible]
44	71	29.1	991	1	SPG_BPK36Q	LYSOGATE N	[illegible]
45	70	28.7	295	1	Y207_SYNY3	HYPOTHETICA	[illegible]

ALIGNMENTS

RESULT 1
ID SERC_BUCAP STANDARD; PRT; 361 AA.
AC P81435;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE PHOSPHOSERINE AMINOTRANSFERASE (EC 2.6.1.52) (PSA
GN SERC.
OS Buchnera aphidicola (subsp. Schizaphis graminum).
OC Bacteria; Proteobacteria; Gamma subdivision; Buch
RN [1]
RP SEQUENCE FROM N.A.
RA Thao M.L., Baumann P.;
RT "Nucleotide sequence of a DNA fragment from Buch
RT (Aphid endosymbiont) containing the genes aspS trp
RT rpsA-himD-tpiA.";
RL Curr. Microbiol. 35:68 [illegible](1997).
CC -!- CATALYTIC ACTIVITY: O-PHOSPHO-L-SERINE + 2-OX
CC 3-PHOSPHONOOXYPYRUVATE + L-GLUTAMATE.
CC -!- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC -!- PATHWAY: REQUIRED BOTH IN MAJOR PHOSPHORYLATED
CC BIOSYNTHESIS AND IN THE BIOSYNTHESIS OF PYRIDO
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARI
CC -!- SIMILARITY: BELONGS TO CLASS V OF PYRIDOXAL PH
CC AMINOTRANSFERASES.
CC ---------- [illegible]
CC This SWISS-PROT entry is copyright. It is produced
CC between the Swiss Institute of Bioinformatics an
CC the European Bioinformatics Institute. There are
CC use by non-profit institutions as long as its
CC modified and this statement is not removed. Usage
CC entities requires a license agreement (See http://
CC or send an email to license@isb-sib.ch).
CC ---------- [illegible]
DR EMBL; L43549; AAC05405.1; -.
DR PFAM; PF00266; aminotran_5; 1.
DR PROSITE; PS00595; AA_TRANSFER_CLASS_5; 1.
KW Serine biosynthesis; Pyridoxine biosynthesis; Tra
KW Aminotransferase; Pyridoxa. phosphate.
FT BINDING 197 197 PYRIDOXAL PHOSPHATE (
SQ SEQUENCE 361 AA; 41260 MW; [illegible]

Query Match 35.7%; Score 87; DB 1; Leng
Best Local Similarity 42.9%; Pred. No. 9.59e-02;
Matches 12; Conservative 7; Mismatches 9; Gaps

Qy 202 SGITILIPKDLIGYALKISPSIFNYY 229

```
          ::: ||: : || ||   ||:::| ||
Qy       3 ASLNTILQQVDLTGYLLVISSTLFGHYI 30


RESULT    2
ID   MCH_RAT        STANDARD;      PRT;   165 AA.
AC   P14200;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   MELANIN-CONCENTRATING HORMONE PRECURSOR.
GN   PMCH OR MCH.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=HYPOTHALAMUS;
RX   MEDLINE; 90005146.
RA   Nahon J.-L., Presse F., Bittencourt J.C., Sawchenko P.E., Vale W.;
RT   "The rat melanin-concentrating hormone messenger ribonucleic acid
RT   encodes multiple putative neuropeptides coexpressed in the
RT   dorsolateral hypothalamus.";
RL   Endocrinology 125:2056-2065(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91083836.
RA   Thompson R.C., Watson S.J.;
RT   "Nucleotide sequence and tissue specific expression of the rat
RT   melanin concentrating hormone gene.";
RL   DNA Cell Biol. 9:637-645(1990).
RN   [3]
RP   SEQUENCE OF 147-165.
RC   TISSUE=HYPOTHALAMUS;
RX   MEDLINE; 89338286.
RA   Vaughan J.M., Fischer W.H., Hoeger C., Rivier J., Vale W.;
RT   "Characterization of melanin-concentrating hormone from rat
RT   hypothalamus.";
RL   Endocrinology 125:1660-1665(1989).
CC   -!- FUNCTION: THIS HORMONE INDUCES MELANOSOMAL AGGREGATION WITHIN
CC       MELANOPHORES AND HAS AN ACTIVITY ANTAGONISTIC TO MELANOTROPIN.
CC   -!- SIMILARITY: BELONGS TO THE MELANIN-CONCENTRATING HORMONE FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M29712; AAA41580.1; -.
DR   EMBL; M62641; AAA41581.1; -.
DR   PIR; A36237; A36237.
DR   PIR; A37407; A37407.
KW   Signal; Cleavage on pair of basic residues; Hormone; Hypothalamus;
KW   Neuropeptide.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22    165       PRO-MCH.
FT   PEPTIDE     109    128       NEUROPEPTIDE NGE (POTENTIAL).
FT   PEPTIDE     131    143       NEUROPEPTIDE NEI (POTENTIAL).
FT   PEPTIDE     147    165       MELANIN-CONCENTRATING HORMONE.
SQ   SEQUENCE   165 AA;  18482 MW;  38E9F07693E77A05 CRC64;

  Query Match           34.0%;  Score 83;  DB 1;  Length 165;
  Best Local Similarity  37.0%;  Pred. No. 3.76e-01;
  Matches   10;  Conservative   11;  Mismatches   5;  Indels   1;  Gaps   1;

Db       1 MAKMSLSSYMLMLAFSLFSHGI-LLSA 26
           :  : |::|:|::: :||:| | ||:|
Qy       9 LQQVDLTGYLLVISSTLFGHYIALLAA 35
```

```
RESULT    3
ID   MCH_HUMAN      STANDARD;      PRT;   165 AA.
AC   P20382; Q16044;
DT   01-FEB-1991 (Rel. 17, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   MELANIN-CONCENTRATING HORMONE PRECURSOR.
GN   PMCH OR MCH.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=HYPOTHALAMUS;
RX   MEDLINE; 91125371.
RA   Presse F., Nahon J.-L., Fischer W.H., Vale W.;
RT   "Structure of the human melanin concentrating hormone mRNA.";
RL   Mol. Endocrinol. 4:632-637(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 93316802.
RA   Breton C., Schorpp M., Nahon J.-L.;
RT   "Isolation and characterization of the human melanin-concentrating
RT   hormone gene and a variant gene.";
RL   Brain Res. Mol. Brain Res. 18:297-310(1993).
CC   -!- FUNCTION: THIS HORMONE INDUCES MELANOSOMAL AGGREGATION WITHIN
CC       MELANOPHORES AND HAS AN ACTIVITY ANTAGONISTIC TO MELANOTROPIN.
CC   -!- SIMILARITY: BELONGS TO THE MELANIN-CONCENTRATING HORMONE FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M57703; AAA63214.1; -.
DR   EMBL; S63697; AAB27493.1; -.
DR   PIR; A34551; A34551.
DR   MIM; 176795; -.
KW   Signal; Cleavage on pair of basic residues; Hormone; Hypothalamus;
KW   Neuropeptide.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22    165       PRO-MCH.
FT   PEPTIDE     109    128       NEUROPEPTIDE NGE (POTENTIAL).
FT   PEPTIDE     131    143       NEUROPEPTIDE NEI (POTENTIAL).
FT   PEPTIDE     147    165       MELANIN-CONCENTRATING HORMONE.
FT   CONFLICT     42     42       P -> S (IN REF. 2).
FT   CONFLICT    104    106       PYL -> GYQ (IN REF. 1).
FT   CONFLICT    113    113       D -> A (IN REF. 2).
SQ   SEQUENCE   165 AA;  18723 MW;  D639E8938B637244 CRC64;

  Query Match           33.6%;  Score 82;  DB 1;  Length 165;
  Best Local Similarity  33.3%;  Pred. No. 5.25e-01;
  Matches    9;  Conservative   13;  Mismatches   4;  Indels   1;  Gaps   1;

Db       1 MAKMNLSSYILILTFSLFSQGI LLSA 26
           :  ::|:::|:|::: :|||: | ||:|
Qy       9 LQQVDLTGYLLVISSTLFGHYIALLAA 35


RESULT    4
ID   MCH_HYLLA      STANDARD;      PRT;    71 AA.
AC   O62690;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   MELANIN-CONCENTRATING HORMONE PRECURSOR (FRAGMENT).
GN   PMCH.
OS   Hylobates lar (Common gibbon).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC   Mammalia; Eutheria; Primates; Catarrhini; Hylobatidae; Hylobates.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98152316.
RA   Viale A., Ortola C., Richard F., Vernier P., Presse F., Schilling S.,
RA   Dutrillaux B., Nahon J.-L.;
RT   "Emergence of a brain-expressed variant melanin concentrating hormone
RT   gene during higher primate evolution: a gene 'in search of a
RT   function'.";
RL   Mol. Biol. Evol. 15:196-214(1998).
CC   -!- FUNCTION: THIS HORMONE INDUCES MELANOSOMAL AGGREGATION WITHIN
CC       MELANOPHORES AND HAS AN ACTIVITY ANTAGONISTIC TO MELANOTROPIN.
CC   -!- SIMILARITY: BELONGS TO THE MELANIN-CONCENTRATING HORMONE FAMILY.
CC   ---------- ---------- ---------- ----------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ---------- ---------- ----------
DR   EMBL; AF029398; AAC05251.1; -.
KW   Signal; Cleavage on pair of basic residues; Hormone; Hypothalamus;
KW   Neuropeptide.
FT   NON_TER       1      1
FT   SIGNAL        <1     20        POTENTIAL.
FT   CHAIN         21     71        PRO-MCH.
FT   NON_TER       71     71
SQ   SEQUENCE   71 AA;  7940 MW;  83D6AD9D8C80BB7F CRC64;

  Query Match             32.4%;  Score 79;  DB 1;  Length 71;
  Best Local Similarity   37.5%;  Pred. No. 1.4e-00;
  Matches     9;  Conservative    12;  Mismatches    2;  Indels    1;  Gaps    1;

Db       3 MNLSSYILILTFSLFSQGI-LLSA 25
           ::|:::|:::: :|||: | ||:|
Qy      12 VDLTGYLLVISSILFGHYIALLAA 35


RESULT    5
ID   MLCH_PANTR     STANDARD;      PRT;    71 AA.
AC   O62689;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   MELANIN-CONCENTRATING HORMONE PRECURSOR (FRAGMENT).
GN   PMCH.
OS   Pan troglodytes (Chimpanzee).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98152316.
RA   Viale A., Ortola C., Richard F., Vernier P., Presse F., Schilling S.,
RA   Dutrillaux B., Nahon J.-L.;
RT   "Emergence of a brain-expressed variant melanin concentrating hormone
RT   gene during higher primate evolution: a gene 'in search of a
RT   function'.";
RL   Mol. Biol. Evol. 15:196-214(1998).
CC   -!- FUNCTION: THIS HORMONE INDUCES MELANOSOMAL AGGREGATION WITHIN
CC       MELANOPHORES AND HAS AN ACTIVITY ANTAGONISTIC TO MELANOTROPIN.
CC   -!- SIMILARITY: BELONGS TO THE MELANIN-CONCENTRATING HORMONE FAMILY.
CC   ---------- ---------- ---------- ----------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ---------- ---------- ----------
DR   EMBL; AF029397; AAC05250.1; -.
KW   Signal; Cleavage on pair of basic residues; Hormone; Hypothalamus;
KW   Neuropeptide.
FT   NON_TER       1      1
FT   SIGNAL        <1     20        POTENTIAL.
FT   CHAIN         21     71        PRO-MCH.
FT   NON_TER       71     71
SQ   SEQUENCE   71 AA;  7942 MW;  CF743E9D9D70BB72 CRC64;

  Query Match             32.4%;  Score 79;  DB 1;  Length 71;
  Best Local Similarity   37.5%;  Pred. No. 1.4e-00;
  Matches     9;  Conservative    12;  Mismatches    2;  Indels    1;  Gaps    1;

Db       3 MNLSSYILILTFSLFSQGI-LLSA 25
           ::|:::|:::: :|||: | ||:|
Qy      12 VDLTGYLLVISSILFGHYIALLAA 35


RESULT    6
ID   MLCH_PANPA     STANDARD;      PRT;    71 AA.
AC   O62688;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   MELANIN-CONCENTRATING HORMONE PRECURSOR (FRAGMENT).
GN   PMCH.
OS   Pan paniscus (Pygmy chimpanzee) (Bonobo).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98152316.
RA   Viale A., Ortola C., Richard F., Vernier P., Presse
RA   Dutrillaux B., Nahon J.-L.;
RT   "Emergence of a brain-expressed variant melanin co
RT   gene during higher primate evolution: a gene 'in s
RT   function'.";
RL   Mol. Biol. Evol. 15:196-214(1998).
CC   -!- FUNCTION: THIS HORMONE INDUCES MELANOSOMAL AGG
CC       MELANOPHORES AND HAS AN ACTIVITY ANTAGONISTIC
CC   -!- SIMILARITY: BELONGS TO THE MELANIN-CONCENTRATI
CC   ---------- ---------- ---------- ----------
CC   This SWISS-PROT entry is copyright. It is produced
CC   between the Swiss Institute of Bioinformatics an
CC   the European Bioinformatics Institute. There are
CC   use by non-profit institutions as long as its
CC   modified and this statement is not removed. Usage
CC   entities requires a license agreement (See http://
CC   or send an email to license@isb-sib.ch).
CC   ---------- ---------- ---------- ----------
DR   EMBL; AF029396; AAC05249.1; -.
KW   Signal; Cleavage on pair of basic residues; Hormone;
KW   Neuropeptide.
FT   NON_TER       1      1
FT   SIGNAL        <1     20        POTENTIAL.
FT   CHAIN         21     71        PRO-MCH.
FT   NON_TER       71     71
SQ   SEQUENCE   71 AA;  7942 MW;  CF743E9D9D70BB72 CRC64;

  Query Match             32.4%;  Score 79;  DB 1;  Length
  Best Local Similarity   37.5%;  Pred. No. 1.4e-00;
  Matches     9;  Conservative    12;  Mismatches    2;

Db       3 MNLSSYILILTFSLFSQGI-LLSA 25
           ::|:::|:::: :|||: | ||:|
Qy      12 VDLTGYLLVISSILFGHYIALLAA 35


RESULT    7
ID   SLFA_PSEAE     STANDARD;      PRT;   186 AA.
AC   O31038;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
```

```
DE   NADH-DEPENDENT FMN REDUCTASE (EC 1.6.8.1).
GN   SLFA.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC   Pseudomonas.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PAO1;
RA   Wuest T., Rist M., Leisinger T., Kertesz M.A.;
RT   "Characterization of the sulfate-regulated sltABC gene cluster of
RT   Pseudomonas aeruginosa.";
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: NADH + FMN = NAD(+) + FMNH(2).
CC   -!- SIMILARITY: BELONGS TO THE SSUE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF026067; AAB82563.1; -.
KW   Oxidoreductase; NAD; FMN.
SQ   SEQUENCE   186 AA;  20028 MW;  752153F9A832ED3C CRC64;

  Query Match           32.4%;  Score 79;  DB 1;  Length 186;
  Best Local Similarity  45.5%;  Pred. No. 1.4le+00;
  Matches   11;  Conservative    8;  Mismatches  12;  Indels   0;  Gaps   0;

Db      62 AELQAVLDEIESADLLLVASPVYRGSYPGLL 92
           | |  :|:::: :  ||| |:   | | :||
Qy       3 ASLNILQQVDLTGYLLVISSTLFGRYIALL 33


RESULT    8
ID   PSBC_CYACA     STANDARD;      PRT;   473 AA.
AC   P28254;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   PHOTOSYSTEM II 44 KDA REACTION CENTER PROTEIN (P6 PROTEIN) (CP43).
GN   PSBC.
OS   Cyanidium caldarium (Galdieria sulphuraria).
OG   Chloroplast.
OC   Eukaryota; Rhodophyta; Bangiophyceae; Porphyridiales; Porphyridiaceae;
OC   Cyanidium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=14-1-1 / ISOLATE 107.79/GOETTINGEN;
RX   MEDLINE; 92216053.
RA   Kessler U., Maid U., Zetsche K.;
RT   "An equivalent to bacterial ompR genes is encoded on the plastid
RT   genome of red algae.";
RL   Plant Mol. Biol. 18:777-780(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=14-1-1 / ISOLATE 107.79/GOETTINGEN;
RX   MEDLINE; 92379254.
RA   Maid U., Zetsche K.;
RT   "A 16 kb small single-copy region separates the plastid DNA inverted
RT   repeat of the unicellular red alga Cyanidium caldarium: physical
RT   mapping of the IR-flanking regions and nucleotide sequences of the
RT   psbD, psbC, rps16, 5S rRNA and rpl23 genes.";
RL   Plant Mol. Biol. 19:1001-1010(1992).
CC   -!- FUNCTION: THE 43 KDA PROTEIN (P6) IS A COMPONENT OF THE CORE OF
CC       PHOTOSYSTEM II. IT IS A CHLOROPHYLL BINDING PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE PSBB / PSBC FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62578; CAA44460.1; -.
DR   PIR; S20856; F2KK4C.
DR   PIR; S25086; S25086.
DR   PFAM; PF00421; PSII; 1.
KW   Photosystem II; Chloroplast; Chlorophyll; Transmembrane.
SQ   SEQUENCE   473 AA;  52487 MW;  7EE7B27D8F712F01 CRC64;

  Query Match           32.4%;  Score 79;  DB 1;  Length 473;
  Best Local Similarity  37.0%;  Pred. No. 1.41e+00;
  Matches   10;  Conservative    7;  Mismatches  10;  Indels   0;  Gaps   0;

Db     157 MTTILGIHLILLGIGAFLLVIKAMFFG 183
           | : | | |  : : ::|||| : :||
Qy       1 MDASLNIILQQVDLTGYLLVISSTLFG 27


RESULT    9
ID   Y841_METTH     STANDARD;      PRT;   343 AA.
AC   O26929;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   HYPOTHETICAL PROTEIN MTH841.
GN   MTH841.
OS   Methanobacterium thermoautotrophicum.
OC   Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC   Methanobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DELTA H;
RX   MEDLINE; 98037514.
RA   Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA   Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA   Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA   Spadafora R., Vicaire R., Wang Y., Wierzbowski J., Gibson R.,
RA   Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA   McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA   Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;
RT   "Complete genome sequence of Methanobacterium thermoautotrophicum
RT   deltaH: functional analysis and comparative genomics.";
RL   J. Bacteriol. 179:7135-7155(1997).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC   -!- SIMILARITY: BELONGS TO THE UPF0077 (PECM) FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000861; AAB85339.1; -.
DR   PFAM; PF03892; DUF6; 2.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM     13     33       POTENTIAL.
FT   TRANSMEM     44     64       POTENTIAL.
FT   TRANSMEM     71     91       POTENTIAL.
FT   TRANSMEM    121    141       POTENTIAL.
FT   TRANSMEM    148    168       POTENTIAL.
FT   TRANSMEM    177    197       POTENTIAL.
FT   TRANSMEM    203    223       POTENTIAL.
FT   TRANSMEM    244    264       POTENTIAL.
FT   TRANSMEM    269    289       POTENTIAL.
FT   TRANSMEM    296    316       POTENTIAL.
FT   TRANSMEM    320    340       POTENTIAL.
SQ   SEQUENCE   343 AA;  37839 MW;  4296C7890CB91B7F CRC64;
```

```
  Query Match           32.0%;  Score 78;  DB 1;  Length 343;
  Best Local Similarity  32.0%;  Pred. No. 1.96e+00;
  Matches    8;  Conservative   12;  Mismatches   5;  Indels   0;  Gaps   0;

Db    193 PGDESILRNVAVICNELVMAAAAFF 217
          :: :: ||: | :|| :|||:::::|
Qy      2 DASLNIILQQVDLIGYLLVISSTLF 26


RESULT   10
ID   TRPA_PORPU     STANDARD;      PRT;   273 AA.
AC   P51382;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   TRYPTOPHAN SYNTHASE ALPHA CHAIN (EC 4.2.1.20).
GN   TRPA.
OS   Porphyra purpurea.
OG   Chloroplast.
OC   Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AVONPORT;
RA   Reith M.E., Munholland J.;
RT   "Complete nucleotide sequence of the Porphyra purpurea chloroplast
RT   genome.";
RL   Plant Mol. Biol. Rep. 13:333-335(1995).
CC   -!- FUNCTION: THE ALPHA SUBUNIT IS RESPONSIBLE FOR THE ALDOL CLEAVAGE
CC       OF INDOLEGLYCEROL PHOSPHATE TO INDOLE AND GLYCERALDEHYDE 3-
CC       PHOSPHATE.
CC   -!- CATALYTIC ACTIVITY: L-SERINE + 1-(INDOL-3-YL)GLYCEROL 3-PHOSPHATE
CC       = L-TRYPTOPHAN + GLYCERALDEHYDE 3-PHOSPHATE + H(2)O.
CC   -!- PATHWAY: LAST (FIFTH) STEP IN BIOSYNTHESIS OF TRYPTOPHAN.
CC   -!- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS
CC       (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE TRPA FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U38804; AAC08268.1; -.
DR   HSSP; P00929; 1BKS.
DR   MENDEL; 10362; PORpu;trpA;1.
DR   PFAM; PF00290; trp_syntA; 1.
DR   PROSITE; PS00167; TRP_SYNTHASE_ALPHA; 1.
KW   Tryptophan biosynthesis; Lyase; Chloroplast.
SQ   SEQUENCE   273 AA;  29835 MW;  2F6C6A072DA8A9AB CRC64;

  Query Match           31.6%;  Score 77;  DB 1;  Length 273;
  Best Local Similarity  36.7%;  Pred. No. 2.70e+00;
  Matches   11;  Conservative    9;  Mismatches  10;  Indels   0;  Gaps   0;

Db     85 NLNNILDMVNIISSNIVAPIVIFIYYNPVL 114
          :|| ||: |:::|  :| :  || | :::|
Qy      4 SLNIILQQVDLTGYLLVISSTLFGHYIALL 33


RESULT   11
ID   OPSB_MOUSE     STANDARD;      PRT;   346 AA.
AC   P51491;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   BLUE-SENSITIVE OPSIN (BLUE CONE PHOTORECEPTOR PIGMENT).
GN   BCP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=RETINA;
RX   MEDLINE; 94375074.
RA   Chiu M.I., Zack D.J., Wang Y., Nathans J.;
RT   "Murine and bovine blue-cone pigment genes: cloning,
RT   characterization of two new members of the S family
RT   pigments.";
RL   Genomics 21:440-443(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Chiu M.I., Zack D.J., Wang Y., Nathans J.;
RL   Submitted (MAR-1997) to the EMBL/GenBank/DDBJ data
CC   -!- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBI
CC       MEDIATE VISION. THEY CONSIST OF AN APOPROTEIN,
CC       LINKED TO CIS-RETINAL.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN
CC   -!- TISSUE SPECIFICITY: THE THREE COLOR PIGMENTS A
CC       PHOTORECEPTOR CELLS.
CC   -!- PTM: SOME OR ALL OF THE C-TERMINAL SER
CC       BE PHOSPHORYLATED.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN C
CC       OPSIN SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced
CC   between the Swiss Institute of Bioinformatics and
CC   the European Bioinformatics Institute. There are n
CC   use by non-profit institutions as long as its
CC   modified and this statement is not removed. Usage
CC   entities requires a license agreement (See http://
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U49720; AAA92458.1; -.
DR   EMBL; U92562; AAB53320.1; -.
DR   EMBL; U92558; AAB53320.1; JOINED.
DR   EMBL; U92559; AAB53320.1; JOINED.
DR   EMBL; U92560; AAB53320.1; JOINED.
DR   EMBL; U92561; AAB53320.1; JOINED.
DR   GCRDB; GCR_1295; -.
DR   GCRDB; GCR_1652; -.
DR   MGD; MGI:99438; BCP.
DR   PFAM; PF00001; 7tm_1; 1.
DR   PFAM; PF00233; 3PSIN.
DR   PRINTS; PR00574; OPSINBLUE.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
DR   PROSITE; PS00238; OPSIN; 1.
KW   Photoreceptor; Retinal protein; Transmembrane; Gly
KW   Phosphorylation; G-protein coupled receptor.
FT   DOMAIN        1      31        EXTRACELLULAR.
FT   TRANSMEM     32      56        1 (POTENTIAL).
FT   DOMAIN       57      68        CYTOPLASMIC.
FT   TRANSMEM     69      94        2 (POTENTIAL).
FT   DOMAIN       95     108        EXTRACELLULAR.
FT   TRANSMEM    109     128        3 (POTENTIAL).
FT   DOMAIN      129     147        CYTOPLASMIC.
FT   TRANSMEM    148     171        4 (POTENTIAL).
FT   DOMAIN      172     197        EXTRACELLULAR.
FT   TRANSMEM    198     225        5 (POTENTIAL).
FT   DOMAIN      226     247        CYTOPLASMIC.
FT   TRANSMEM    248     271        6 (POTENTIAL).
FT   DOMAIN      272     279        EXTRACELLULAR.
FT   TRANSMEM    280     304        7 (POTENTIAL).
FT   DOMAIN      305     346        CYTOPLASMIC.
FT   CARBOHYD     12      12        PROBABLE.
FT   DISULFID    105     182        POTENTIAL.
FT   BINDING     291     291        RETINAL CHROMOPHORE.
SQ   SEQUENCE   346 AA;  38921 MW;  8BEEAB0198EE6864 CRC

  Query Match           31.6%;  Score 77;  DB 1;  Length
  Best Local Similarity  45.4%;  Pred. No. 2.70e+00;
  Matches   13;  Conservative    5;  Mismatches   8;  Indels

Db     66 PLNYILVNVSLGGFLICIFS VFIVFIA 92
```

```
          :|| ||  | | |:|: | | :|  :||
Qy       4 SLNIILQQVDLTGYLLVISSTLFGHYIA 31


RESULT   12
ID   OPSB_RAT       STANDARD;      PRT;   346 AA.
AC   Q63652;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   BLUE-SENSITIVE OPSIN (BLUE CONE PHOTORECEPTOR PIGMENT).
GN   BCP.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE DAWLEY; TISSUE=PINEAL GLAND;
RX   MEDLINE; 97227593.
RA   Zhao X., Haeseleer F., Fariss R.N., Huang J., Baehr W., Milam A.H.,
RA   Palczewski K.;
RT   "Molecular cloning and localization of rhodopsin kinase in the
RT   mammalian pineal.";
RL   Vis. Neurosci. 14:225-232(1997).
CC   -!- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
CC       MEDIATE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY
CC       LINKED TO CIS-RETINAL.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: THE THREE COLOR PIGMENTS ARE FOUND IN THE CONE
CC       PHOTORECEPTOR CELLS.
CC   -!- PTM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY
CC       BE PHOSPHORYLATED.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC       OPSIN SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U63972; AAB05931.1; -.
DR   GCRDB; GCR_1410; -.
DR   PFAM; PF00001; 7tm_1; 1.
DR   PRINTS; PR00238; OPSIN.
DR   PRINTS; PR00574; OPSINBLUE.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
DR   PROSITE; PS00238; OPSIN; 1.
KW   Photoreceptor; Retinal protein; Transmembrane; Glycoprotein; Vision;
KW   Phosphorylation; G-protein coupled receptor.
FT   DOMAIN        1     31       EXTRACELLULAR.
FT   TRANSMEM     32     56       1 (POTENTIAL).
FT   DOMAIN       57     68       CYTOPLASMIC.
FT   TRANSMEM     69     94       2 (POTENTIAL).
FT   DOMAIN       95    108       EXTRACELLULAR.
FT   TRANSMEM    109    128       3 (POTENTIAL).
FT   DOMAIN      129    147       CYTOPLASMIC.
FT   TRANSMEM    148    171       4 (POTENTIAL).
FT   DOMAIN      172    197       EXTRACELLULAR.
FT   TRANSMEM    198    225       5 (POTENTIAL).
FT   DOMAIN      226    247       CYTOPLASMIC.
FT   TRANSMEM    248    271       6 (POTENTIAL).
FT   DOMAIN      272    279       EXTRACELLULAR.
FT   TRANSMEM    280    304       7 (POTENTIAL).
FT   DOMAIN      305    346       CYTOPLASMIC.
FT   CARBOHYD     12     12       PROBABLE.
FT   DISULFID    105    182       POTENTIAL.
FT   BINDING     291    291       RETINAL CHROMOPHORE.
SQ   SEQUENCE   346 AA;  39057 MW;  1B5CA2273219FF9E CRC64;

  Query Match           31.6%;  Score 77;  DB 1;  Length 346;
  Best Local Similarity  46.4%;  Pred. No. 2.70e+00;
  Matches   13;  Conservative    5;  Mismatches   9;  Indels   1;  Gaps   1;

Db      66 PLNYILVNVSLGGFLFCIFS-VFTVFIA 92
           :|| ||  | | |:|: | | :|  :||
Qy       4 SLNIILQQVDLTGYLLVISSTLFGHYIA 31


RESULT   13
ID   FXSA_SERMA     STANDARD;      PRT;   139 AA.
AC   P47148;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   FXSA PROTEIN (FRAGMENT).
GN   FXSA.
OS   Serratia marcescens.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Serratia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SR41;
RX   MEDLINE; 95148727.
RA   Omori K., Akatsuka H., Komatsubara S.;
RT   "Construction of a versatile promoter analysis vector and its use for
RT   analysis of the Serratia marcescens aspartase promoter region.";
RL   Plasmid 32:233-247(1994).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC       (POTENTIAL).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D13252; BAA02517.1; -.
DR   PIR; PS0396; PS0396.
KW   Transferase; Inner membrane.
FT   TRANSMEM     17     37       POTENTIAL.
FT   TRANSMEM     78     98       POTENTIAL.
FT   NON_TER     139    139
SQ   SEQUENCE   139 AA;  14996 MW;  A897780614A5A09D CRC64;

  Query Match           31.1%;  Score 76;  DB 1;  Length 139;
  Best Local Similarity  40.0%;  Pred. No. 3.71e+00;
  Matches    8;  Conservative    9;  Mismatches   2;  Indels   1;  Gaps   1;

Db      80 LAGFLLLIPG-FFTDFLGLL 98
           |:|:||:|:: :|  :::||
Qy      14 LTGYLLVISSTLFGHYIALL 33


RESULT   14
ID   ROLB_AGRRB     STANDARD;      PRT;   262 AA.
AC   P49409;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   PROTEIN-TYROSINE PHOSPHATASE ROLB (EC 3.1.3.48) (ROL B PROTEIN).
GN   ROLB.
OS   Agrobacterium rhizogenes.
OG   Plasmid pRil724.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Agrobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MAFF03 01724;
RX   MEDLINE; 94211915.
RA   Piyokawa S., Kobayashi Y., Kikuchi Y., Kamada H., Harada H.;
RT   "Root inducing region of mikimopine type Ri plasmid pRil724.";
```

```
RL   Plant Physiol. 104:801-802(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MAFF03-01724;
RX   MEDLINE; 94227335.
RA   Tanaka N., Ikeda T., Oka A.;
RT   "Nucleotide sequence of the rol region of the mikimopine-type root-
RT   inducing plasmid pRi1724.";
RL   Biosci. Biotechnol. Biochem. 58:548-551(1994).
CC   -!- FUNCTION: INDUCES DIFFERENTIATION AND GROWTH OF NEOPLASTIC ROOTS
CC       (HAIRY ROOTS). SEEMS TO FUNCTION AS A TYROSINE PHOSPHATASE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X64255; CAA45540.1; -.
DR   EMBL; AB006689; BAA22335.1; ALT_INIT.
KW   T-DNA; Plasmid; Hydrolase.
SQ   SEQUENCE   262 AA;  30097 MW;  32DB34DCDA200BF4 CRC64;

  Query Match             41.1%;  Score 76;  DB 1;  Length 262;
  Best Local Similarity   37.5%;  Pred. No. 3.7le+00;
  Matches     9;  Conservative    9;  Mismatches   6;  Indels   0;  Gaps   0;

Db     155 VEEPDLEGFIAIFASILFIHLLEL 178
           ::: || |:: :::|||| | : |
Qy       9 LQQVDLTGYLLVISSTLFGHYIAL 32


RESULT   15
ID   CITG_ECOLI     STANDARD;      PRT;   292 AA.
AC   P77231;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   CITG PROTEIN.
GN   CITG.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE; 97426617.
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner F.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,
RA   Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H., Lin D.,
RA   Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PROBABLY INVOLVED IN THE BIOSYNTHESIS OF THE PROSTHETIC
CC       GROUP (2'(5''-PHOSPHORIBOSYL) 3' DEPHOSPHOCOENZYME-A) OF THE GAMMA
CC       SUBUNIT OF CITRATE LYASE.
CC   -!- SIMILARITY: BELONGS TO THE CITG FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000166; AAC73714.1;
DR   EMBL; U82598; AAB40814.1;
DR   ECOGENE; EG13539; CITG.
DR   PFAM; PF01874; CitG; 1
SQ   SEQUENCE   292 AA;  31944 MW;  BE3C24BE44E4AE6 CRC64;

  Query Match             41.1%;  Score 76;  DB 1;  Length
  Best Local Similarity   32.6%;  Pred. No. 3.7le+00;
  Matches    10;  Conservative    5;  Mismatches   8;

Db     185 AGYPLVINHAL-PHYLTLL 207
           :|| ||| :|  |||:|.
Qy      15 TGYLLVISSTLFGHYIALL 33


Search completed: Fri Aug 25 16:48:31 2000
Job time : 12 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 25 18:00:20 2000 MasPar time 4.87 Seconds
439.322 Million cell updates/sec

Tabular output not generated.

Title: >US-09-257-585-6
Description: (1-35) from US09257585.pep
Perfect Score: 244
Sequence: 1 MDASLNITLQQVDLIGYLLVISSILFGHYIALLAA 35

Scoring table: PAM 150
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir64
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 34.941; Variance 71.929; scale 0.486

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	87	35.7	305	2	G70382	phosphotidylglycerolph	4.78e-01
2	85	34.8	350	2	I04749	probable iron transpo	8.82e-01
3	85	34.8	418	2	S51260	probable O-antigen tr	8.82e-01
4	83	34.0	165	2	A36237	melanin-concentrating	1.62e+00
5	82	33.6	98	2	T11083	NADH dehydrogenase (u	2.18e+00
6	82	33.6	165	2	I52634	melanin-concentrating	2.18e+00
7	82	33.6	245	2	D29835	ORF protein Escheri	2.18e+00
8	81	33.2	386	2	P00589	hypothetical protein	2.93e+00
9	81	33.2	469	2	G69267	conserved hypothetica	2.93e+00
10	81	33.2	503	2	S31940	starch-degrading enzy	2.93e+00
11	79	32.4	473	1	F2KK4C	photosystem II chloro	5.28e+00
12	78	32.0	74	2	JC4096	hypothetical 8.6K pro	7.05e+00
13	78	32.0	210	2	S65529	NADH dehydrogenase (u	7.05e+00
14	78	32.0	343	2	G69212	conserved hypothetica	7.05e+00
15	77	31.6	273	2	S73303	tryptophan synthase (	9.41e+00
16	77	31.6	346	1	B54679	opsin, blue-sensitive	9.41e+00
17	76	31.1	107	2	S74020	hypothetical protein	1.25e+01
18	76	31.1	139	2	PS0396	hypothetical protein	1.25e+01
19	76	31.1	279	2	S52639	hypothetical protein	1.25e+01
20	76	31.1	279	2	JC2106	rolB protein - Agroba	1.25e+01
21	76	31.1	292	2	G64795	citG protein - Escher	1.25e+01
22	76	31.1	527	2	S26037	NADH dehydrogenase (u	1.25e+01
23	76	31.1	804	2	S61395	probable Na+/H+-excha	1.25e+01
24	75	30.7	176	2	G70158	purine bin	1.65e+01
25	75	30.7	416	2	I13511	hypothetic	1.65e+01
26	75	30.7	486	2	G70579	probable g	1.65e+01
27	75	30.7	488	1	KXHC	coagulatio	1.65e+01
28	75	30.7	522	2	S28733	NADH dehyd	1.65e+01
29	74	30.3	98	1	T10994	NADH dehyd	2.15e+01
30	74	30.3	204	2	I40761	hypothetic	2.15e+01
31	74	30.3	348	1	OPHBUS	opsin, blu	2.15e+01
32	74	30.3	354	2	C72215	peroxidase	2.15e+01
33	74	30.3	474	2	S34746	photosyste	2.15e+01
34	74	30.3	474	2	JN0346	photosyste	2.15e+01
35	74	30.3	835	2	F72624	probable c	2.15e+01
36	73	29.9	140	2	S57063	probable p	2.84e+01
37	73	29.9	333	2	S43789	probable i	2.84e+01
38	73	29.9	394	2	S77242	hypothetic	2.84e+01
39	73	29.9	473	1	F2N144	photosyste	2.84e+01
40	73	29.9	773	2	S77543	hypothetic	2.84e+01
41	72	29.5	98	2	T11382	NADH dehyd	3.74e+01
42	72	29.5	98	2	T11251	NADH dehyd	3.74e+01
43	72	29.5	414	2	A54823	cAMP recep	3.74e+01
44	72	29.5	475	2	S58657	photosyste	3.74e+01
45	72	29.5	1027	1	I45029	cruzi pro	3.74e+01

ALIGNMENTS

RESULT 1
ENTRY G70382 #type complete
TITLE phosphotidylglycerophosphate synthase
ORGANISM #formal_name Aquifex aeolicus
DATE 08-May-1998 #sequence_revision 08-May-1998
ACCESSIONS G70382
REFERENCE A70300
#authors Deckert, G.; Warren, P.V.; Gaasterland, Lenox, A.L.; Graham, D.E.; Overbeek Keller, M.; Aniay, M.; Huber, R.; P J.M.; Olsen, G.J.; Swanson, R.V.
#journal Nature (1998) 392:353-358
#title The complete genome of the hyperthermo Aquifex aeolicus.
#cross-references MUID:98179669
#accession G70382
##status preliminary; nucleic acid sequen translation not shown
##molecule_type DNA
##residues 1-305 ##label AQF
##cross-references GB:AE000774; NID:g2983463; PI
##experimental_source strain VF5
GENETICS
#gene pgsA2
SUMMARY #length 305 #molecular-weight 34785

Query Match 35.7%; Score 87; DB 2; Len
Best Local Similarity 34.5%; Pred. No. 4.78e-01;
Matches 10; Conservative 12; Mismatches 8;

Db 125 SATSLILLGVELGNTLILA VTLA TIS 151
Qy 4 ASLNITLQQVDLIGYLLVISSILFGHYIA 41

RESULT 2
ENTRY I04749 #type complete
TITLE probable iron transport protein [imp thaliana
ORGANISM #formal_name Arabidopsis thaliana #cu cress
DATE 23-Apr-1999 #sequence_revision 23-Apr 11-Jun-1999
ACCESSIONS I04749
REFERENCE 215483
#authors Bevan, M.; De Bruijn, M.; Maarse, A.;

```
                  Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Schueller, C.
   #submission    submitted to the Protein Sequence Database, June 1998
   #accession     T04749
      ##molecule_type DNA
      ##residues       1-350 ##label BEV
      ##cross-references EMBL:AL024486
      ##experimental_source cultivar Columbia; BAC clone T16H5
GENETICS
   #map_position 4
   #introns       199/3; 249/3
   #note          T16H5.40
SUMMARY           #length 350  #molecular-weight 37407  #checksum 1227

  Query Match             34.8%;  Score 85;  DB 2;  Length 350;
  Best Local Similarity   46.7%;  Pred. No. 8.82e-01;
  Matches    11;  Conservative    10;  Mismatches    9;  Indels    0;  Gaps    0;

Db      45 PLKIVAIVAILTTSLIGVTSPLFSRYISFL 74
           :|:|:    ||  |: ::| |||:||::|
Qy       4 SLNIILQQVDLTGYLIVISSTLFGHYIALL 33


RESULT    3
ENTRY             S51260      #type complete
TITLE             probable O-antigen transport protein trsA - Yersinia
                    enterocolitica
ALTERNATE_NAMES   trgA protein
ORGANISM          #formal_name Yersinia enterocolitica
DATE              07-May-1995 #sequence_revision 01-Sep-1995 #text_change
                    24-Sep-1999
ACCESSIONS        S70736; S30030; S51260
REFERENCE         S70734
   #authors       Skurnik, M.; Venho, R.; Toivanen, P.; Al-Hendy, A.
   #journal       Mol. Microbiol. (1995) 17:575-594
   #title         A novel locus of Yersinia enterocolitica serotype O:3
                    involved in lipopolysaccharide outer core biosynthesis.
   #cross-references MUID:96100456
   #accession     S70736
      ##status         nucleic acid sequence not shown; translation not shown
      ##molecule_type DNA
      ##residues       1-418 ##label SKU
      ##cross-references EMBL:Z47767; NID:g633689; PIDN:CAA87698.1;
                          PID:g633692
      ##experimental_source strain 6471/76 serotype O:3
      ##note           the nucleotide sequence was submitted to the EMBL Data
                        Library, January 1995
GENETICS
   #gene          trsA
CLASSIFICATION    #superfamily O-antigen transporter protein
KEYWORDS          lipopolysaccharide core biosynthesis; transmembrane protein
SUMMARY           #length 418  #molecular-weight 47170  #checksum 5837

  Query Match             34.8%;  Score 85;  DB 2;  Length 418;
  Best Local Similarity   66.7%;  Pred. No. 8.82e-01;
  Matches    12;  Conservative     4;  Mismatches    1;  Indels    1;  Gaps    1;

Db     333 YLLVIQSTVLGNYI-LLP 349
           ||||| ||::|:|| ||:
Qy      17 YLLVISSILFGHYIALLA 34


RESULT    4
ENTRY             A36237      #type complete
TITLE             melanin-concentrating hormone precursor - rat
ORGANISM          #formal_name Rattus norvegicus #common_name Norway rat
DATE              12 Apr 1991 #sequence_revision 12 Apr 1991 #text_change
                    16-Jul-1999
ACCESSIONS        A36237; A37407; A37406
REFERENCE         A36237
   #authors       Thompson, R.C.; Watson, S.J.
   #journal       DNA Cell Biol. (1990) 9:637-645
   #title         Nucleotide sequence and tissue-specific expression of the rat
                    melanin concentrating hormone gene.
   #cross-references MUID:91083836
   #accession     A36237
      ##status         preliminary; not compared with conceptual translation
      ##molecule_type DNA
      ##residues       1-165 ##label THO
      ##cross-references GB:M62641; NID:g205329; PIDN:AAA41581.1; PID:g205330
REFERENCE         A37407
   #authors       Nahon, J.L.; Presse, F.; Bittencourt, J.C.; Sawchenko, P.E.;
                    Vale, W.
   #journal       Endocrinology (1989) 125:2056-2065
   #title         The rat melanin-concentrating hormone messenger ribonucleic
                    acid encodes multiple putative neuropeptides coexpressed in
                    the dorsolateral hypothalamus.
   #cross-references MUID:90005146
   #accession     A37407
      ##status         preliminary
      ##molecule_type mRNA
      ##residues       1-165 ##label NAH
      ##cross-references GB:M29712; NID:g205327; PIDN:AAA41580.1; PID:g205328
REFERENCE         A37406
   #authors       Vaughan, J.M.; Fischer, W.H.; Hoeger, C.; Rivier, J.; Vale,
                    W.
   #journal       Endocrinology (1989) 125:1660-1665
   #title         Characterization of melanin concentrating hormone from rat
                    hypothalamus.
   #cross-references MUID:89338286
   #accession     A37406
      ##status         preliminary
      ##molecule_type protein
      ##residues       147-165 ##label VAU
CLASSIFICATION    #superfamily melanin-concentrating hormone
SUMMARY           #length 165  #molecular-weight 18482  #checksum 2551

  Query Match             34.0%;  Score 83;  DB 2;  Length 165;
  Best Local Similarity   37.0%;  Pred. No. 1.62e+00;
  Matches    10;  Conservative    11;  Mismatches    5;  Indels    1;  Gaps    1;

Db       1 MAKMSLSSYMLMLAFSLFSHGI-LLSA 26
           :  : |::|:|::: :|||:| | ||:|
Qy       9 LQQVDLTGYLIVISSTLFGHYIALLAA 35


RESULT    5
ENTRY             T11083      #type complete
TITLE             NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 4L - Falco
                    peregrinus mitochondrion (SGC1)
ORGANISM          #formal_name mitochondrion Falco peregrinus
DATE              16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change
                    20-Sep-1999
ACCESSIONS        T11083
REFERENCE         217242
   #authors       Mindell, D.P.; Sorenson, M.D.; Dimcheff, D.E.
   #journal       Proc. Natl. Acad. Sci. U.S.A. (1998) 95:10693-10697
   #title         Multiple independent origins of mitochondrial gene order in
                    birds.
   #accession     T11083
      ##status         preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues       1-98 ##label MIN
      ##cross-references EMBL:AF090338; NID:g4894462; PID:g4894471;
                          PIDN:AAD32500.1
GENETICS
   #genome        mitochondrion
   #genetic_code SGC1
CLASSIFICATION    #superfamily NADH dehydrogenase (ubiquinone) chain 4L
KEYWORDS          mitochondrion, NAD, oxidoreductase
SUMMARY           #length 98  #molecular-weight 10478  #checksum 1970

  Query Match             33.6%;  Score 82;  DB 2;  Length 98;
  Best Local Similarity   36.7%;  Pred. No. 2.18e+00;
  Matches    11;  Conservative     9;  Mismatches   10;  Indels    0;  Gaps    0;

Db      15 SSLGLTLHRTHLISALLCLESMMLSMYIAL 44
```

```
          :|| : |::  | : || : | ::: ||||
Qy       3 ASLNIILQQVDLTGYLLVISSTLFGHYIAL 32


RESULT    6
ENTRY             I52634      #type complete
TITLE             melanin-concentrating hormone precursor - human
ALTERNATE_NAMES   prepro-melanin-concentrating hormone
ORGANISM          #formal_name Homo sapiens #common_name man
DATE              02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
                    16-Jul-1999
ACCESSIONS        I52634; I65414; A34551
REFERENCE         I52634
   #authors       Breton, C.; Schorpp, M.; Nahon, J.L.
   #journal       Brain Res. Mol. Brain Res. (1993) 18:297-310
   #title         Isolation and characterization of the human
                    melanin-concentrating hormone gene and a variant gene.
   #cross-references MUID:93316802
   #accession     I52634
      ##status         preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues       1-165 ##label RES
      ##cross-references GB:S63697; NID:g402795; PIDN:AAB27493.1; PID:g402796
   #accession     I65414
      ##status         preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues       88-96,'S',98-100,'V',102-108,'E',110-121,'D',123-131,
                        'I',133-142,'V',144-149,'I',151,'S',153-159,'QS',
                        162-165 ##label RE2
      ##cross-references GB:S64288; NID:g402797; PIDN:AAB27494.1; PID:g402798
REFERENCE         A34551
   #authors       Presse, F.; Nahon, J.L.; Fischer, W.H.; Vale, W.
   #journal       Mol. Endocrinol. (1990) 4:632-637
   #title         Structure of the human melanin concentrating hormone mRNA.
   #cross-references MUID:91125371
   #accession     A34551
      ##status         preliminary
      ##molecule_type mRNA
      ##residues       1-41,'F',43-103,'G',105,'Q',107-112,'D',114-165 ##label
                        PRE
      ##cross-references GB:M57703; NID:g187445; PIDN:AAA63214.1; PID:g187446
GENETICS
   #gene          GDB:PMCH
      ##cross-references GDB:128784; OMIM:176795
   #map_position 12q23-12q24
   #introns       83/3; 150/1
CLASSIFICATION    #superfamily melanin-concentrating hormone
KEYWORDS          hormone; hypothalamus
FEATURE
   1-21               #domain signal sequence #status predicted #label SIG
SUMMARY           #length 165  #molecular-weight 18619  #checksum 1939

  Query Match             33.6%;  Score 82;  DB 2;  Length 165;
  Best Local Similarity   38.3%;  Pred. No. 2.18e+00;
  Matches    9;  Conservative   13;  Mismatches   4;  Indels   1;  Gaps   1;

Db       1 MAKMNLSSYILILIFSLFSQGI-LISA 26
           :  :|:|:|:::: :|||: | ||:|
Qy       9 LQQVDLTGYLLVISSILFGHYIALLAA 35


RESULT    7
ENTRY             D29835      #type fragment
TITLE             ORF protein - Escherichia coli plasmid pED208 (fragment)
ORGANISM          #formal_name Escherichia coli
DATE              04-Aug-1988 #sequence_revision 04-Aug-1988 #text_change
                    23-Mar-1993
ACCESSIONS        D29835
REFERENCE         A29835
   #authors       Finlay, B.B.; Paranchych, W.
   #journal       J. Bacteriol. (1986) 166:713-721
   #title         Nucleotide sequence of the surface exclusion genes traS and
                    traT from the IncF-0 lac plasmid pED208.
   #cross-references MUID:86224784
   #accession     D29835
      ##status         preliminary
      ##molecule_type DNA
      ##residues       1-245 ##label FIN
GENETICS
   #genome        plasmid
SUMMARY           #length 245  #checksum 6678

  Query Match             33.6%;  Score 82;  DB 2;  Length
  Best Local Similarity   41.0%;  Pred. No. 2.18e+00;
  Matches    9;  Conservative   12;  Mismatches   8;

Db       3 LKMFLQINNLISYYVIMGIVLPSIAVILM 31
           |:::||  :| :| :::: |:  : :|
Qy       5 LNIILQQVDLTGYLLVISSILFGHYIALL 33


RESULT    8
ENTRY             T00589      #type complete
TITLE             hypothetical protein T27E13.17 - Arab
ORGANISM          #formal_name Arabidopsis thaliana #com
                    cress
DATE              01-Feb-1999 #sequence_revision 01-Feb
                    30-Apr 1999
ACCESSIONS        T00589
REFERENCE         Z34177
   #authors       Bowmsley, S.D.; ...n, X.; Ketchum, K.A.
                    Brandon, R.; Sykes, S.M.; Kark, S.
                    Berlavage, A.E.; Adams, M.D.; Somers
                    J.C.
   #submission    submitted to the EMBL Data Library, M
   #description   Arabidopsis thaliana chromosome II BA
                    sequence.
   #accession     T00589
      ##status         translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues       1-986 ##label ROD
      ##cross-references EMBL:A[illegible]41-5; NID:g4150496; [illegible]
      ##experimental_source cultivar Columbia
GENETICS
   #map_position 2
   #introns       65/3; 182/3
   #note          T27E13.17
SUMMARY           #length 988  #[illegible]ular weight 4[illegible]

  Query Match             33.2%;  Score 81;  DB 2;  Length
  Best Local Similarity   42.8%;  Pred. No. 2.93e+00;
  Matches   10;  Conservative   13;  Mismatches   9;

Db     243 PELKIERAPLS-YIII:KVSEI:PNF:SIII 272
           ::: ::: :|: |:|: : :::::|:
Qy       5 LNIILQQVDLIGYLLVIS SILFGHYIALLA 34


RESULT    9
ENTRY             G69267      #type complete
TITLE             conserved hypothetical protein AF0343
                    fulgidus
ORGANISM          #formal_name Archaeoglobus fulgidus
DATE              05-Dec-1997 #sequence_revision 05-Dec
                    05-Jun 1998
ACCESSIONS        G69267
REFERENCE         A69250
   #authors       Klenk, H.P.; Clayton, R.A.; Tomb, J.F.
                    K.E.; Ketchum, K.A.; Dodson, R.J.; Fu
                    Peterson, J.D.; Richardson, D.L.; Ke
                    D.E.; Kyrpides, N.C.; Fleischmann, R
                    Lee, N.H.; Sutton, G.G.; Gill, S.; K
                    Dougherty, B.A.; McKenny, K.; Adams,
                    Peterson, S.; Reich, C.I.; McNeil, L
                    Glodek, A.; Zhou, L.; Overbeek, R.;
                    J.F., McDonald, L.; Utterback, T.; K
```

```
                  T.; Artiach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.;
                  D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.;
                  Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese,
                  C.R.; Venter, J.C.
   #journal       Nature (1997) 390:364-370
   #title         The complete genome sequence of the hyperthermophilic,
                    sulfate-reducing archaeon Archaeoglobus fulgidus.
   #cross-references MUID:98049343
   #accession     G69267
      ##status          preliminary; nucleic acid sequence not shown;
                         translation not shown
      ##molecule_type DNA
      ##residues       1-469 ##label KLE
      ##cross-references GB:AE001096; GB:AE000782; NID:g2689419; PID:g2650499;
                         TIGR:AF0143
SUMMARY           #length 469  #molecular-weight 52627  #checksum 7785

  Query Match             33.2%;  Score 81;  DB 2;  Length 469;
  Best Local Similarity  38.1%;  Pred. No. 2.93e+00;
  Matches    8;  Conservative    8;  Mismatches   5;  Indels   0;  Gaps   0;

Db      31 FLLTAINLSVFLLIIFTGLFG 51
           ::|  ::|: :||:| : |||
Qy       7 IILQQVDLTGYLLVISSTLFG 27


RESULT   10
ENTRY             S31940      #type complete
TITLE             starch-degrading enzyme   Clostridium acetobutylicum
ORGANISM          #formal_name Clostridium acetobutylicum
DATE              06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change
                    09-Sep-1997
ACCESSIONS        S31940
REFERENCE         S31940
   #authors       Verhasselt, P.; Vanderleyden, J.; Leuven, F.
   #submission    submitted to the EMBL Data Library, February 1993
   #accession     S31940
      ##status          preliminary
      ##molecule_type DNA
      ##residues       1-503 ##label VER
      ##cross-references EMBL:X70334; NID:g49196; PID:g49197
SUMMARY           #length 503  #molecular-weight 53930  #checksum 9332

  Query Match             33.2%;  Score 81;  DB 2;  Length 503;
  Best Local Similarity  38.5%;  Pred. No. 2.98e+00;
  Matches   10;  Conservative    8;  Mismatches   8;  Indels   0;  Gaps   0;

Db       1 MKKIFTNVLLVGFLVFISAILYGHSV 26
           :: |:  | | |:|: ||: |:|| :
Qy       5 LNTILQQVDLTGYLLVISSTLFGHYI 30


RESULT   11
ENTRY             F2KK4C      #type complete
TITLE             photosystem II chlorophyll a-binding protein psbC precursor -
                    red alga (Cyanidium caldarium) chloroplast
ORGANISM          #formal_name chloroplast Cyanidium caldarium
DATE              31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change
                    16-Jul-1999
ACCESSIONS        S20856; S25086
REFERENCE         S20854
   #authors       Kessler, U., Maid, U., Zetsche, K.
   #journal       Plant Mol. Biol. (1992) 18:777-780
   #title         An equivalent to bacterial ompR genes is encoded on the
                    plastid genome of red algae.
   #cross references MUID:92216053
   #accession     S20856
      ##status          translation not shown
      ##molecule_type DNA
      ##residues       1-473 ##label KES
      ##cross-references EMBL:X62578; NID:g11278; PIDN:CAA44460.1; PID:g11281
REFERENCE         S25084
   #authors       Maid, U.; Zetsche, K.
   #journal       Plant Mol. Biol. (1992) 19:1001-1010
   #title         A 16 kb small single copy region separates the plastid DNA
                    inverted repeat of the unicellular red alga Cyanidium
                    caldarium: physical mapping of the IR-flanking regions and
                    nucleotide sequences of the psbD-psbC, rps16, 5S rRNA and
                    rpl21 genes.
   #cross-references MUID:92379254
   #accession     S25086
      ##molecule_type DNA
      ##residues       1-473 ##label MAI
      ##cross-references EMBL:X62578; NID:g11278; PIDN:CAA44460.1; PID:g11281
GENETICS
   #gene          psbC
   #genome        chloroplast
CLASSIFICATION    #superfamily photosystem II chlorophyll a-binding protein
                    psbC
KEYWORDS          chlorophyll; chloroplast; photosynthesis; photosystem II;
                    thylakoid
FEATURE
   1-14               #domain transit peptide (thylakoid) #status predicted
                        #label TRP\
   15-473             #product photosystem II chlorophyll a-binding protein
                        psbC #status predicted #label MAT
SUMMARY           #length 473  #molecular-weight 52487  #checksum 4379

  Query Match             32.4%;  Score 79;  DB 1;  Length 473;
  Best Local Similarity  37.0%;  Pred. No. 5.28e+00;
  Matches   10;  Conservative    7;  Mismatches  10;  Indels   0;  Gaps   0;

Db     157 MTTILGIHLILLGIGAFLLVIKAMFFG 183
           |.|| | |  : : ::|||| . .||
Qy       1 MDASLNIILQQVDLTGYLLVISSTLFG 27


RESULT   12
ENTRY             JC4096      #type complete
TITLE             hypothetical 8.6K protein   Cyanobacterium synechococcus
ORGANISM          #formal_name Cyanobacterium synechococcus
DATE              16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change
                    16-Jul-1999
ACCESSIONS        JC4096
REFERENCE         JC4094
   #authors       Sugita, M.; Luo, L.; Ohta, M.; Itadani, H.; Matsubayashi, T.;
                    Sugiura, M.
   #journal       DNA Res. (1995) 2:71-76
   #title         Genes encoding the group I intron-containing tRNALeu and
                    subunit L of NADH dehydrogenase from the Cyanobacterium
                    synechococcus PCC 6301
   #accession     JC4096
      ##status          preliminary
      ##molecule_type DNA
      ##residues       1-74 ##label SUG
      ##cross-references DDBJ:D42186
SUMMARY           #length 74  #molecular-weight 8561  #checksum 8661

  Query Match             32.0%;  Score 78;  DB 2;  Length 74;
  Best Local Similarity  37.0%;  Pred. No. 7.05e+00;
  Matches   10;  Conservative    8;  Mismatches   8;  Indels   1;  Gaps   1;

Db       1 MTVTLIIAALYLALAGAYLLVVPAALY 27
           |  .|.|    . |:| |||....|:
Qy       1 MDASLNIILQQVDLTG-YLLVISSTLF 26


RESULT   13
ENTRY             S65529      #type complete
TITLE             NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) nqrD protein -
                    Vibrio alginolyticus
ALTERNATE_NAMES   NADH-ubiquinone oxidoreductase nqrD protein
ORGANISM          #formal_name Vibrio alginolyticus
DATE              17-Dec-1996 #sequence_revision 13-Mar-1997 #text_change
                    29-Sep-1999
ACCESSIONS        S65529; S51018
```

```
REFERENCE       S65529
   #authors     Hayashi, M.; Hirai, K.; Unemoto, T.
   #journal     FEBS Lett. (1995) 363:75-77
   #title       Sequencing and the alignment of structural genes in the nqr
                  operon encoding the Na(+)-translocating NADH-quinone
                  reductase from Vibrio alginolyticus.
   #cross-references MUID:95246889
   #accession   S65529
      ##status        preliminary
      ##molecule_type DNA
      ##residues      1-210 ##label HAY
      ##cross-references EMBL:D49364; NID:g893412; PID:d1008557; PID:g893414
REFERENCE       S51018
   #authors     Beattie, P.; Tan, K.; Bourne, R.M.; Leach, D.; Rich, P.R.;
                  Ward, F.B.
   #journal     FEBS Lett. (1994) 356:333-338
   #title       Cloning and sequencing of four structural genes for the Na
                  (+)-translocating NADH-ubiquinone oxidoreductase of Vibrio
                  alginolyticus.
   #cross-references MUID:95104445
   #accession   S51018
      ##status        preliminary
      ##molecule_type DNA
      ##residues      1-210 ##label BEA
      ##cross-references EMBL:Z37111; NID:g663268; PIDN:CAA85479.1;
                        PID:g663273
GENETICS
   #gene        nqr 4; nqrD
CLASSIFICATION  #superfamily conserved hypothetical protein HI1688
KEYWORDS        NAD; oxidoreductase
SUMMARY         #length 210  #molecular weight 22602  #checksum 9319

  Query Match           32.0%;  Score 78;  DB 2;  Length 210;
  Best Local Similarity 39.4%;  Pred. No. 7.05e+00;
  Matches   13;  Conservative   10;  Mismatches   8;  Indels   2;  Gaps   2;

Db      80 ASLVIVVDQV-LKAYLYDISKQL-SVFVGLIIT 110
           ||| |:::|| | ::| || .. : :::|: :
Qy       3 ASLNIILQQVDLTGYLLVISSILFGHYIALLAA 35


RESULT   14
ENTRY           C69212      #type complete
TITLE           conserved hypothetical protein MTH841 - Methanobacterium
                  thermoautotrophicum (strain Delta H)
ORGANISM        #formal_name Methanobacterium thermoautotrophicum
DATE            05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
                  05-Jun-1998
ACCESSIONS      C69212
REFERENCE       A69000
   #authors     Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.;
                  Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakely, D.;
                  Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.;
                  Lumm, W.; Pothier, B.; Qiu, D.; Spadafora, R.; Vicaire, R.;
                  Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Caruso,
                  A.; Bush, D.; Safer, H.; Patwell, D.; Prabhakar, S.;
                  McDougall, S.; Shimer, G.; Goyal, A.; Pietrokovski, S.;
                  Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling,
                  J.; Reeve, J.N.
   #journal     J. Bacteriol. (1997) 179:7135-7155
   #title       Complete genome sequence of Methanobacterium
                  thermoautotrophicum Delta H: functional analysis and
                  comparative genomics.
   #cross-references MUID:98047514
   #accession   C69212
      ##status        preliminary; nucleic acid sequence not shown;
                       translation not shown
      ##molecule_type DNA
      ##residues      1-343 ##label MTH
      ##cross-references GB:AE000861; GB:AE000666; NID:g2621940; PID:g2621932
      ##experimental_source strain Delta H
GENETICS
   #gene        MTH841
SUMMARY         #length 343  #molecular weight 37859

  Query Match           32.0%;  Score 78;  DB 2;  Length
  Best Local Similarity 32.0%;  Pred. No. 7.05e+00;
  Matches    8;  Conservative   12;  Mismatches   [illegible]

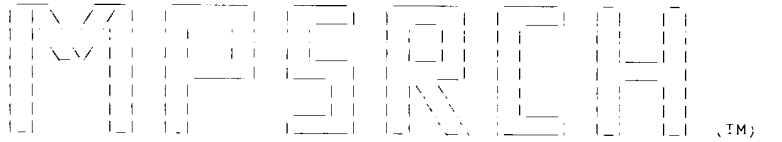
Db     193 EKDFSTILRNVAVIGNIVMAAAFF 217
           :: :: ||: | :|| :|:::::|
Qy       2 DASLNIILQQVDLTGYLLVISSILF 26


RESULT   15
ENTRY           S73303      #type complete
TITLE           tryptophan synthase (EC 4.2.1.20) alph
                  (Porphyra purpurea) chloroplast
ORGANISM        #formal_name chloroplast Porphyra purp
DATE            19-Mar-1997 #sequence_revision 09-May
                  16-Jul 1999
ACCESSIONS      S73303
REFERENCE       S73108
   #authors     Reith, M.; Munholland, J.
   #journal     Plant Mol. Biol. Rep. (1995) 13:333-3
   #title       Complete nucleotide sequence of the Po
                  chloroplast genome.
   #accession   S73303
      ##status        nucleic acid sequence not shown;
      ##molecule_type DNA
      ##residues      1-273 ##label REI
      ##cross-references EMBL:U[illegible]4; NID:g1276652; PID:
                        PID:g1276[illegible]
      ##note          the nucleotide sequence was submi
                       library, October 1995
GENETICS
   #gene        trpA
   #genome      chloroplast
CLASSIFICATION  #superfamily tryptophan synthase alpha
                  synthase alpha-chain homology
KEYWORDS        carbon-oxygen lyase; chloroplast; tryp
                  biosynthesis
FEATURE
   26-253           #domain tryptophan synthase alpha
                      IREA
SUMMARY         #length 273  #molecular weight 29[illegible]

  Query Match           31.6%;  Score 77;  DB 2;  Leng
  Best Local Similarity 36.7%;  Pred. No. 9.41e+00;
  Matches   11;  Conservative    9;  Mismatches  10;

Db      85 NLNNILDMVNIISSNIVAEIVEIYYNVL 114
           :|| ||: |:::|: : : :  : |:::|
Qy       4 SLNIILQQVDLTGYLLVISSILFGHYIALL 33


Search completed: Fri Aug 25 18:[illegible]4m 2000
Job time : 28 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 25 18:02:55 2000; MasPar time 2.30 Seconds
233.214 Million cell updates/sec

Tabular output not generated.

Title: >US-09-257-585-6
Description: (1-345) from US09257585.pep
Perfect Score: 244
Sequence: 1 MDASLNIILQQVDLIGYLLVISSTLFGHYIALLAA 345

Scoring table: PAM 150
Gap 11

Searched: 152433 seqs, 15429240 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:SA_COMB 2:SB_COMB 3:6_COMB 4:PCI_COMB 5:backfilesl

Statistics: Mean 21.995; Variance 101.904; scale 0.216

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	85	34.8	345	2	US-08-758-	Sequence 14, Applicati	6.57e+00
2	75	30.7	437	1	US-08-487-	Sequence 2, Applicatio	3.67e+01
3	75	30.7	437	1	US-08-487-	Sequence 3, Applicatio	3.67e+01
4	75	30.7	488	1	US-08-487	Sequence 1, Applicatio	3.67e+01
5	74	29.9	140	1	US-08-691-	Sequence 4, Applicatio	5.14e+01
6	74	29.9	140	2	US-08-843-	Sequence 4, Applicatio	5.14e+01
7	72	29.5	346	5	5474933-2	atent No. 5474933	6.07e+01
8	72	29.5	346	5	5474933-7	atent No. 5474933	6.07e+01
9	72	29.5	346	2	US-08-476-	Sequence 10, Applicati	6.07e+01
10	72	29.5	346	2	US-08-476-	Sequence 2, Applicatio	6.07e+01
11	72	29.5	834	2	US-08-396-	Sequence 4, Applicatio	6.07e+01
12	72	29.5	834	2	US-08-861-	Sequence 4, Applicatio	6.07e+01
13	71	29.1	2,000	2	US-08-477-	Sequence 8, Applicatio	7.13e+01
14	70	28.7	332	1	US-08-469-	Sequence 2, Applicatio	8.48e+01
15	70	28.7	766	2	US-08-951-	Sequence 2, Applicatio	8.48e+01
16	70	28.7	779	2	US-08-951-	Sequence 6, Applicatio	8.48e+01
17	70	28.7	803	2	US-08-951-	Sequence 4, Applicatio	8.48e+01
18	68	27.9	191	1	US-08-389-	Sequence 4, Applicatio	1.18e+02
19	68	27.9	191	4	PCT-US95-0	Sequence 4, Applicatio	1.18e+02
20	68	27.9	191	1	US-08-732-	Sequence 4, Applicatio	1.18e+02
21	68	27.9	592	2	US-08-646-	Sequence 30, Applicati	1.18e+02
22	68	27.9	592	3	US-09-069	Sequence 30, Applicati	1.18e+02
23	68	27.9	592	2	US-08-599-	Sequence 30, Applicati	1.18e+02
24	67	27.5	346	1	US-08-118	Sequence 58,	[illegible]
25	67	27.5	346	4	PCT-US93-0	Sequence 50,	[illegible]
26	67	27.5	771	2	US-08-928	Sequence 26,	[illegible]
27	67	27.5	811	1	US-07-937	Sequence 21,	[illegible]
28	67	27.5	836	2	US-08-846-	Sequence 95,	[illegible]
29	67	27.5	1727	2	US-08-477	Sequence 10,	[illegible]
30	66	27.0	525	1	US-08-752-	Sequence 19,	[illegible]
31	66	27.0	525	1	US-08-375	Sequence 19,	[illegible]
32	65	26.6	405	1	US-08-118-	Sequence 99,	[illegible]
33	65	26.6	462	4	PCT-US93-0	Sequence 59,	[illegible]
34	65	26.6	462	2	US-08-466-	Sequence 10,	[illegible]
35	65	26.6	492	1	US-08-911-	Sequence 35,	[illegible]
36	65	26.6	980	2	US-08-822-	Sequence 2,	[illegible]
37	65	26.6	980	1	US-08-203-	Sequence 2,	[illegible]
38	65	26.6	980	1	US-08-091-	Sequence 2,	[illegible]
39	65	26.6	1027	2	US-08-551-	Sequence 2,	[illegible]
40	65	26.6	1027	1	US-09-004-	Sequence 2,	[illegible]
41	64	26.2	320	2	US-08-600-	Sequence 4,	[illegible]
42	64	26.2	378	2	US-09-055-	Sequence 1,	[illegible]
43	64	26.2	1477	1	US-08-345	Sequence 3,	[illegible]
44	64	26.2	1477	1	US-08-492-	Sequence 10,	[illegible]
45	64	26.2	1477	1	US-08-423-	Sequence 10,	[illegible]

ALIGNMENTS

RESULT 1
ID US-08-758-621-14 STANDARD; PRT; 345 AA.
XX
AC vvvvvv
XX
DT
XX
DE Sequence 14, Applicat[illegible] US/08758621
XX
CC Sequence 14, Application US/08758621
CC Patent No. 5846821
CC GENERAL INFORMATION:
CC APPLICANT: Guerinot, Mary Lou, and Eide, Dav[illegible]
CC TITLE OF INVENTION: Metal-Regulated Transport[illegible]
CC NUMBER OF SEQUENCES: 21
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: LAHIVE & COCKFIELD
CC STREET: 60 State Street
CC CITY: Boston
CC STATE: Massachusetts
CC COUNTRY: USA
CC ZIP: 02109-1875
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/758,621
CC FILING DATE:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/08/670,578
CC FILING DATE: 25-MAY-1996
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Silveri, Joan M.
CC REGISTRATION NUMBER: 39,558
CC REFERENCE/DOCKET NUMBER: INCO 099CP
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617)227-7400
CC TELEFAX: (617)227-5640
CC INFORMATION FOR SEQ ID NO: 14:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 345 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 345 AA; 36758 MW; 34164 CN;

```
  Query Match             24.8%;  Score 85;  DB 2;  Length 345;
  Best Local Similarity   36.7%;  Pred. No. 6.57e+00;
  Matches   11;  Conservative   10;  Mismatches   9;  Indels   0;  Gaps   0;

Db      40 PLKIVAIVAILFTSLIGVTSPLFSRYISFL 69
           :|:|:      ||  |: ::| ||::||::|
Qy       4 SLNIILQQVDLIGYLLVISSILFGHYIALL 33


RESULT    2
ID   US-08-487-037-2     STANDARD;     PRT;   437 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application US/08487037
XX
CC    Sequence 2, Application US/08487037
CC    Patent No. 5795863
CC     GENERAL INFORMATION:
CC       APPLICANT:  Wolf, David L.
CC       TITLE OF INVENTION:  RECOMBINANT AGENTS AFFECTING THROMBOSIS
CC       NUMBER OF SEQUENCES:  11
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  MORRISON & FOERSTER
CC         STREET:  2000 Pennsylvania Avenue, NW
CC         CITY:  Washington
CC         STATE:  DC
CC         COUNTRY:  USA
CC         ZIP:  20006-1812
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/487,037
CC         FILING DATE:  07-JUN 1995
CC         CLASSIFICATION:  514
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Adler, Reid G.
CC         REGISTRATION NUMBER:  30,988
CC         REFERENCE/DOCKET NUMBER:  2802-0002.02
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (202) 887-1500
CC         TELEFAX:  (202) 887 0763
CC         TELEX:  90 4030
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  437 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  both
CC       FEATURE:
CC         NAME/KEY:  Protein
CC         LOCATION:  1..139
CC         OTHER INFORMATION:  /note= "Factor Xa Light Chain"
CC       FEATURE:
CC         NAME/KEY:  Peptide
CC         LOCATION:  -40..0
CC         OTHER INFORMATION:  /note= "Pre-Pro leader sequence"
CC       FEATURE:
CC         NAME/KEY:  Modified-site
CC         LOCATION:  -17
CC         OTHER INFORMATION:  /note= "Location of Intron A"
CC       FEATURE:
CC         NAME/KEY:  Modified-site
CC         LOCATION:  (37^38)
CC         OTHER INFORMATION:  /note= "Location of Intron B"
CC       FEATURE:
CC         NAME/KEY:  Modified-site
CC         LOCATION:  46
CC         OTHER INFORMATION:  /note= "Location of Intron C"
CC       FEATURE:
CC         NAME/KEY:  Modified-site
CC         LOCATION:  63
CC         OTHER INFORMATION:  /note= "An amino acid represented
CC         OTHER INFORMATION:  by the greek letter Beta"
CC       FEATURE:
CC         NAME/KEY:  Modified-site
CC         LOCATION:  84
CC         OTHER INFORMATION:  /note= "Location of Intron D"
CC       FEATURE:
CC         NAME/KEY:  Modified-site
CC         LOCATION:  128
CC         OTHER INFORMATION:  /note= "Location of Intron E"
CC       FEATURE:
CC         NAME/KEY:  Modified-site
CC         LOCATION:  (158^159)
CC         OTHER INFORMATION:  /note= "Location of Intron F"
CC       FEATURE:
CC         NAME/KEY:  Modified-site
CC         LOCATION:  198
CC         OTHER INFORMATION:  /note= "Location of Intron G"
CC       FEATURE:
CC         NAME/KEY:  Disulfide-bond
CC         LOCATION:  group(17..22, 50..61, 55..70, 72..81, 89..100, 96
CC         LOCATION:  ..109, 111..124, 132..251, 150..155, 170..186,
CC         LOCATION:  255..313, 324..452)
SQ   SEQUENCE   437 AA; 48895 MW; 995777 CN;

  Query Match             30.7%;  Score 75;  DB 1;  Length 437;
  Best Local Similarity   34.6%;  Pred. No. 3.67e+01;
  Matches    9;  Conservative    9;  Mismatches   7;  Indels   1;  Gaps   1;

Db       1 MGRPLHLVLLSASLAG-LLLLGESLF 25
           |  :|:::|    |:| ||::: :||
Qy       1 MDASLNIILQQVDETGYLLVISSTLF 26


RESULT    3
ID   US-08-487-037-3     STANDARD;     PRT;   437 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 3, Application US/08487037
XX
CC    Sequence 3, Application US/08487037
CC    Patent No. 5795863
CC     GENERAL INFORMATION:
CC       APPLICANT:  Wolf, David L.
CC       TITLE OF INVENTION:  RECOMBINANT AGENTS AFFECTING THROMBOSIS
CC       NUMBER OF SEQUENCES:  11
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  MORRISON & FOERSTER
CC         STREET:  2000 Pennsylvania Avenue, NW
CC         CITY:  Washington
CC         STATE:  DC
CC         COUNTRY:  USA
CC         ZIP:  20006-1812
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/487,037
CC         FILING DATE:  07-JUN-1995
CC         CLASSIFICATION:  514
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Adler, Reid G.
CC         REGISTRATION NUMBER:  30,988
```

```
CC         REFERENCE/DOCKET NUMBER:  2803-0002.02
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (202) 887-1500
CC         TELEFAX:  (202) 887-0763
CC         TELEX:  90-4030
CC     INFORMATION FOR SEQ ID NO:  3:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  437 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  both
CC       FEATURE:
CC         NAME/KEY:  Modified-site
CC         LOCATION:  -40..397
CC         OTHER INFORMATION:  /note= "Some features apply from
CC         OTHER INFORMATION:  SEQ ID NO:2"
CC       FEATURE:
CC         NAME/KEY:  Protein
CC         LOCATION:  1..139
CC         OTHER INFORMATION:  /note= "Factor Xa - Light Chain"
CC       FEATURE:
CC         NAME/KEY:  Peptide
CC         LOCATION:  -40..0
CC         OTHER INFORMATION:  /note= "Pre-Pro leader sequence"
CC       FEATURE:
CC         NAME/KEY:  Modified-site
CC         LOCATION:  -17
CC         OTHER INFORMATION:  /note= "Location of Intron A"
CC       FEATURE:
CC         NAME/KEY:  Modified-site
CC         LOCATION:  (37'38)
CC         OTHER INFORMATION:  /note= "Location of Intron B"
CC       FEATURE:
CC         NAME/KEY:  Modified-site
CC         LOCATION:  46
CC         OTHER INFORMATION:  /note= "Location of Intron C"
CC       FEATURE:
CC         NAME/KEY:  Modified-site
CC         LOCATION:  63
CC         OTHER INFORMATION:  /note= "An amino acid represented
CC         OTHER INFORMATION:  by the greek letter Beta"
CC       FEATURE:
CC         NAME/KEY:  Modified-site
CC         LOCATION:  84
CC         OTHER INFORMATION:  /note= "Location of Intron D"
CC       FEATURE:
CC         NAME/KEY:  Modified site
CC         LOCATION:  (158'159)
CC         OTHER INFORMATION:  /note= "Location of Intron D"
CC       FEATURE:
CC         NAME/KEY:  Modified-site
CC         LOCATION:  198
CC         OTHER INFORMATION:  /note= "Location of Intron C"
CC       FEATURE:
CC         NAME/KEY:  Disulfide-bond
CC         LOCATION:  group(17..22, 50..61, 55..70, 72..81, 89..100, 96
CC         LOCATION:  ..109, 111..124, 132..251, 150..155, 170..186,
CC         LOCATION:  299..313, 324..352)
SQ   SEQUENCE   437 AA; 48878 MW; 989986 CN;

  Query Match           30.7%;  Score 75;  DB 1;  Length 437;
  Best Local Similarity  34.6%;  Pred. No. 3.67e+01;
  Matches    9;  Conservative    9;  Mismatches   7;  Indels   1;  Gaps   1;

Db       1 MGRPLHLVLLSASLAG-LLLLGESLF 25
           |  :|:::|   |:| ||::: :||
Qy       1 MDASLNIILQQVDLIGYLLVISSILF 26


RESULT    4
ID   US-08-487-037-1     STANDARD;     PRT;   488 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 1, Application US/08487037
XX
CC    Sequence 1, Application US/08487037
CC    Patent No. 5795863
CC     GENERAL INFORMATION:
CC       APPLICANT:  Wolf, David C.
CC       TITLE OF INVENTION:  RECOMBINANT AGENTS AFFEC
CC       NUMBER OF SEQUENCES:  11
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  MORRISON & FOERSTER
CC         STREET:  2000 Pennsylvania Avenue, NW
CC         CITY:  Washington
CC         STATE:  DC
CC         COUNTRY:  USA
CC         ZIP:  20006-1812
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #.
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/487,037
CC         FILING DATE:  07-JUN-1995
CC         CLASSIFICATION:  514
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Adler, Reid S.
CC         REGISTRATION NUMBER:  30,988
CC         REFERENCE/DOCKET NUMBER:  2803-0002.02
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (202) 887-1500
CC         TELEFAX:  (202) 887-0763
CC         TELEX:  90-4030
CC     INFORMATION FOR SEQ ID NO:  1:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  488 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  both
CC       FEATURE:
CC         NAME/KEY:  Modified site
CC         LOCATION:  -17
CC         OTHER INFORMATION:  /note= "Location of Int
CC       FEATURE:
CC         NAME/KEY:  Modified site
CC         LOCATION:  (37'38)
CC         OTHER INFORMATION:  /note= "Location of Int
CC       FEATURE:
CC         NAME/KEY:  Modified site
CC         LOCATION:  46
CC         OTHER INFORMATION:  /note= "Location of Int
CC       FEATURE:
CC         NAME/KEY:  Modified-site
CC         LOCATION:  63
CC         OTHER INFORMATION:  /note= "Amino acid repr
CC         OTHER INFORMATION:  the greek letter Beta"
CC       FEATURE:
CC         NAME/KEY:  Modified site
CC         LOCATION:  84
CC         OTHER INFORMATION:  /note= "Location of Int
CC       FEATURE:
CC         NAME/KEY:  Modified-site
CC         LOCATION:  128
CC         OTHER INFORMATION:  /note= "Location of Int
CC       FEATURE:
CC         NAME/KEY:  Modified site
CC         LOCATION:  (209'210)
CC         OTHER INFORMATION:  /note= "Location of Int
CC       FEATURE:
CC         NAME/KEY:  Modified site
CC         LOCATION:  249
```

```
CC         OTHER INFORMATION:  /note= "Location of Intron 6"
CC       FEATURE:
CC         NAME/KEY:  Peptide
CC         LOCATION:  -40..0
CC         OTHER INFORMATION:  /note= "Pre-Pro leader sequence"
CC       FEATURE:
CC         NAME/KEY:  Protein
CC         LOCATION:  1..139
CC         OTHER INFORMATION:  /note= "Factor Xa- Light chain"
CC       FEATURE:
CC         NAME/KEY:  Peptide
CC         LOCATION:  143..194
CC         OTHER INFORMATION:  /note= "Activation Peptide"
CC       FEATURE:
CC         NAME/KEY:  Protein
CC         LOCATION:  195..448
CC         OTHER INFORMATION:  /note= "Factor Xa Heavy Chain"
CC       FEATURE:
CC         NAME/KEY:  Disulfide-bond
CC         LOCATION:  group(17..22, 50..61, 55..70, 72..81, 89..100, 96
CC         LOCATION:  ..109, 111..124, 132..192, 201..206, 221..237,
CC         LOCATION:  350..364, 375..403)
SQ   SEQUENCE   488 AA; 54418 MW; 1242431 CN;

  Query Match           30.7%;  Score 75;  DB 1;  Length 488;
  Best Local Similarity  34.6%;  Pred. No. 3.67e+01;
  Matches    9;  Conservative    9;  Mismatches   7;  Indels   1;  Gaps   1;

Db       1 MGRPLHLVLLSASLAG-LLLLGESLF 25
           |  :|::::|    |:| ||:::  :||
Qy       1 MDASLNIILQQVDLTGYLLVISSTLF 26


RESULT    5
ID   US-08-691-071-4     STANDARD;     PRT;   140 AA.
XX
AC   XXXXXX
XX
DT
XX
DE   Sequence 4, Application US/08691071
XX
CC    Sequence 4, Application US/08691071
CC    Patent No. 5789198
CC     GENERAL INFORMATION:
CC       APPLICANT:  Akerblom, Ingrid E.
CC       TITLE OF INVENTION:  A NOVEL HUMAN LEPTIN
CC       TITLE OF INVENTION:  RECEPTOR-RELATED PROTEIN
CC       NUMBER OF SEQUENCES:  4
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Incyte Pharmaceuticals, Inc.
CC         STREET:  3174 Porter Drive
CC         CITY:  Palo Alto
CC         STATE:  CA
CC         COUNTRY:  U.S.
CC         ZIP:  94304
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette
CC         COMPUTER:  IBM Compatible
CC         OPERATING SYSTEM:  DOS
CC         SOFTWARE:  FastSEQ Version 1.5
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/691,071
CC         FILING DATE:  Filed Herewith
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Billings, Lucy J.
CC         REGISTRATION NUMBER:  36,749
CC         REFERENCE/DOCKET NUMBER:  PF-0111 US
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  415-855-0555
CC         TELEFAX:  415-845-4166
CC     INFORMATION FOR SEQ ID NO:  4:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  140 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  peptide
CC       IMMEDIATE SOURCE:
CC         LIBRARY:  GenBank
CC         CLONE:  1197072
SQ   SEQUENCE   140 AA; 15663 MW; 109452 CN;

  Query Match           29.9%;  Score 73;  DB 1;  Length 140;
  Best Local Similarity  52.6%;  Pred. No. 5.14e+01;
  Matches   10;  Conservative    5;  Mismatches   3;  Indels   1;  Gaps   1;

Db      22 GFLLVTLSCALFHNYYPLF 40
           |:|||| | :|| :| :|:
Qy      16 GYLLVI SSTLFGHYIALL 33


RESULT    6
ID   US-08-843-370-4     STANDARD;     PRT;   140 AA.
XX
AC   XXXXXX
XX
DT
XX
DE   Sequence 4, Application US/08843370
XX
CC    Sequence 4, Application US/08843370
CC    Patent No. 5874535
CC     GENERAL INFORMATION:
CC       APPLICANT:  Akerblom, Ingrid E.
CC       TITLE OF INVENTION:  A NOVEL HUMAN LEPTIN RECEPTOR
CC       TITLE OF INVENTION:  GENE-RELATED PROTEIN
CC       NUMBER OF SEQUENCES:  4
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Incyte Pharmaceuticals, Inc.
CC         STREET:  3174 Porter Drive
CC         CITY:  Palo Alto
CC         STATE:  CA
CC         COUNTRY:  U.S.
CC         ZIP:  94304
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette
CC         COMPUTER:  IBM Compatible
CC         OPERATING SYSTEM:  DOS
CC         SOFTWARE:  FastSEQ Version 1.5
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/843,370
CC         FILING DATE:  Filed Herewith
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 08/691,071
CC         FILING DATE:  August 1, 1996
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Billings, Lucy J.
CC         REGISTRATION NUMBER:  36,749
CC         REFERENCE/DOCKET NUMBER:  PF-0111-1 US
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  415-855-0555
CC         TELEFAX:  415-845-4166
CC     INFORMATION FOR SEQ ID NO:  4:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  140 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  peptide
CC       IMMEDIATE SOURCE:
CC         LIBRARY:  GenBank
CC         CLONE:  1197072
SQ   SEQUENCE   140 AA; 15663 MW; 109452 CN;

  Query Match           29.9%;  Score 73;  DB 2;  Length 140;
```

```
  Best Local Similarity  52.6%;  Pred. No. 5.14e+01;
  Matches    10;  Conservative    5;  Mismatches   3;  Indels   1;  Gaps   1;

Db      22 GFLLVILSCALFHNYYPLF 40
           |:|||| | :|| :| :|:
Qy      16 GYLLVI-SSILFGHYIALL 33


RESULT    7
ID   5474933-2      STANDARD;      PRT;   346 AA.
XX
AC   xxxxxx
XX
DT
XX
DE  Patent No. 5474933
XX
CC   Patent No. 5474933
CC        APPLICANT: EINER, RONALD M.;FUQUA JR., WILLIAM C.
CC        TITLE OF INVENTION: MARINE MELA GENE
CC        NUMBER OF SEQUENCES: 9
CC        CURRENT APPLICATION DATA:
CC          APPLICATION NUMBER:  US/08/148,945
CC          FILING DATE: 08-NOV-1993
CC        PRIOR APPLICATION DATA:
CC          APPLICATION NUMBER: 974,837
CC          FILING DATE: 10-NOV-1992
CC          APPLICATION NUMBER: 496,804
CC          FILING DATE: 21-MAR-1990
CC   SEQ ID NO:2:
CC          LENGTH: 346
SQ   SEQUENCE   346 AA; 39453 MW; 585300 CN;

  Query Match           29.5%;  Score 72;  DB 5;  Length 346;
  Best Local Similarity  39.4%;  Pred. No. 6.07e+01;
  Matches    11;  Conservative    7;  Mismatches   8;  Indels   2;  Gaps   2;

Db     289 ILVDGDEDGYLLQIFTKNLFGPIFIEII 316
           || : |  |||| :::  ||| :| ::
Qy       8 ILQQVDLIGYLL-VISSTLFGH-YIALL 33


RESULT    8
ID   5474933-7      STANDARD;      PRT;   346 AA.
XX
AC   xxxxxx
XX
DT
XX
DE  Patent No. 5474933
XX
CC   Patent No. 5474933
CC        APPLICANT: EINER, RONALD M.;FUQUA JR., WILLIAM C.
CC        TITLE OF INVENTION: MARINE MELA GENE
CC        NUMBER OF SEQUENCES: 9
CC        CURRENT APPLICATION DATA:
CC          APPLICATION NUMBER:  US/08/148,945
CC          FILING DATE: 08-NOV-1993
CC        PRIOR APPLICATION DATA:
CC          APPLICATION NUMBER: 974,837
CC          FILING DATE: 10-NOV-1992
CC          APPLICATION NUMBER: 496,804
CC          FILING DATE: 21-MAR-1990
CC   SEQ ID NO:7:
CC          LENGTH: 346
SQ   SEQUENCE   346 AA; 39475 MW; 586455 CN;

  Query Match           29.5%;  Score 72;  DB 5;  Length 346;
  Best Local Similarity  39.4%;  Pred. No. 6.07e+01;
  Matches    11;  Conservative    7;  Mismatches   8;  Indels   2;  Gaps   2;

Db     289 ILVDGDEDGYLLQIFTKNLFGPIFIEII 316
           || : |  |||| :::  ||| :| ::
Qy       8 ILQQVDLIGYLL-VISSTLFGH-YIALL 33


RESULT    9
ID   US-08-476-254-10      STANDARD;      PRT;   346 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 10, Application US/08476254
XX
CC    Sequence 10, Application US/08476254
CC    Patent No. 5846531
CC     GENERAL INFORMATION:
CC       APPLICANT:  WEINER, RONALD M.
CC       APPLICANT:  FUQUA, WILLIAM C.
CC       TITLE OF INVENTION:  MARINE MELA GENE
CC       NUMBER OF SEQUENCES:  11
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  WATSON COLE STEVENS DAVIS, P.L.L.
CC         STREET:  1400 K STREET NW
CC         CITY:  WASHINGTON
CC         STATE:  D.C.
CC         COUNTRY:  USA
CC         ZIP:  20005-2477
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/476,254
CC         FILING DATE:  06-JUN-1995
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  POULOS III, JAMES A.
CC         REGISTRATION NUMBER:  31,724
CC         REFERENCE/DOCKET NUMBER:  IAP403100
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  202 628 0808
CC         TELEFAX:  202 628 8 31
CC     INFORMATION FOR SEQ ID NO:  1 :
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  346 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  peptide
CC       HYPOTHETICAL:  NO
CC       ANTI-SENSE:  NO
CC       ORIGINAL SOURCE:
CC         ORGANISM:  S. colwelliana
CC         STRAIN:  D
SQ   SEQUENCE   346 AA; 39453 MW; 585300 CN;

  Query Match           29.5%;  Score 72;  DB 2;  Length
  Best Local Similarity  39.4%;  Pred. No. 6.07e+01;
  Matches    11;  Conservative    7;  Mismatches   8;

Db     289 ILVDGDEDGYLLQIFTKNLFGPIFIEII 316
           || : |  |||| :::  ||| :| ::
Qy       8 ILQQVDLIGYLL-VISSTLFGH-YIALL 33


RESULT   10
ID   US-08-476-254-2      STANDARD;      PRT;   346 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application US/08476254
```

```
XX
CC    Sequence 2, Application US/08476254
CC    Patent No. 5846531
CC     GENERAL INFORMATION:
CC       APPLICANT:  WEINER, RONALD M.
CC       APPLICANT:  FUQUA, WILLIAM C.
CC       TITLE OF INVENTION:  MARINE MELA GENE
CC       NUMBER OF SEQUENCES:  11
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  WATSON COLE STEVENS DAVIS, P.L.L.C.
CC         STREET:  1400 K STREET NW
CC         CITY:  WASHINGTON
CC         STATE:  D.C.
CC         COUNTRY:  USA
CC         ZIP:  20005-2477
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/476,254
CC         FILING DATE:  06-JUN-1995
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  POULOS III, JAMES A.
CC         REGISTRATION NUMBER:  31,714
CC         REFERENCE/DOCKET NUMBER:  JAP303196
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  202 628-0088
CC         TELEFAX:  202 628-8044
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  346 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   346 AA; 39483 MW; 586736 CN;

  Query Match           29.5%;  Score 72;  DB 2;  Length 346;
  Best Local Similarity  39.3%;  Pred. No. 6.07e+01;
  Matches   11;  Conservative    7;  Mismatches   8;  Indels   2;  Gaps   2;

Db    289 ILVDGDEDGYLLQIFTKNLFGPIFIEII 316
          || : |  |||| :::  ||| : | ::
Qy      8 ILQQVDLTCYLL-VISSTLFGH-YIALL 33


RESULT   11
ID   US-08-396-001-4      STANDARD;      PRT;   834 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 4, Application US/08396001
XX
CC    Sequence 4, Application US/08396001
CC    Patent No. 5919618
CC     GENERAL INFORMATION:
CC       APPLICANT:  Guarente, Leonard P.
CC       APPLICANT:  Austriaco Jr., Nicanor
CC       APPLICANT:  Claus, James
CC       APPLICANT:  Cole, Francesca
CC       APPLICANT:  Kennedy, Brian
CC       TITLE OF INVENTION:  Genes Determining Cellular Senesence in
CC       TITLE OF INVENTION:  Yeast
CC       NUMBER OF SEQUENCES:  16
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
CC         STREET:  Two Militia Drive
CC         CITY:  Lexington
CC         STATE:  MA
CC         COUNTRY:  USA
CC         ZIP:  02173
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/396,001
CC         FILING DATE:  28-FEB-1995
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Granahan, Patricia
CC         REGISTRATION NUMBER:  32,227
CC         REFERENCE/DOCKET NUMBER:  MIT-6408A2
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  617-861-6240
CC         TELEFAX:  617-861-9540
CC     INFORMATION FOR SEQ ID NO:  4:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  834 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   834 AA; 92828 MW; 3480242 CN;

  Query Match           29.5%;  Score 72;  DB 2;  Length 834;
  Best Local Similarity  42.4%;  Pred. No. 6.07e+01;
  Matches   14;  Conservative    8;  Mismatches   9;  Indels   2;  Gaps   2;

Db    466 INASMNILLTTIDIFTVNLNVLIRDNPGNY-AL 497
          ::||:|||:|  :|: |  | |:    ||:| ||
Qy      1 MDASLNIILQQVDL-TGYLLVISSTLFGHYIAL 32


RESULT   12
ID   US-08-861-464-4      STANDARD;      PRT:   834 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 4, Application US/08861464
XX
CC    Sequence 4, Application US/08861464
CC    Patent No. 5874210
CC     GENERAL INFORMATION:
CC       APPLICANT:  Guarente, Leonard P.
CC       APPLICANT:  Austriaco Jr., Nicanor
CC       APPLICANT:  Kennedy, Brian
CC       TITLE OF INVENTION:  Genes Determining Cellular Senescence
CC       TITLE OF INVENTION:  in Yeast
CC       NUMBER OF SEQUENCES:  16
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
CC         STREET:  Two Militia Drive
CC         CITY:  Lexington
CC         STATE:  MA
CC         COUNTRY:  USA
CC         ZIP:  02173
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/861,464
CC         FILING DATE:  22-MAY-1997
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 08/396,001
CC         FILING DATE:  28-FEB-1995
CC       PRIOR APPLICATION DATA:
```

```
CC         APPLICATION NUMBER:  PCT/US94/09351
CC         FILING DATE:  15-AUG-1994
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 08/107,408
CC         FILING DATE:  16-AUG-1993
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Granahan, Patricia
CC         REGISTRATION NUMBER:  32,227
CC         REFERENCE/DOCKET NUMBER:  MIT-6408A22
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  781-861-6240
CC         TELEFAX:  781-861-9540
CC     INFORMATION FOR SEQ ID NO:  4:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  834 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   834 AA; 92828 MW; 3480242 CN;

  Query Match             29.5%;  Score 72;  DB 2;  Length 834;
  Best Local Similarity  42.4%;  Pred. No. 6.07e+01;
  Matches   14;  Conservative    9;  Mismatches   9;  Indels   2;  Gaps   2;

Db    466 INASMNILETTIDIFTVNLNVLIRDNFCNY-AL 497
          ::|:|:||  :|: | | |:    |||: |:
Qy      1 MDASLNIILQQVDL-TGYLLVISSILFGHYIAL 32


RESULT   13
ID   US-08-477-451-8      STANDARD;      PRT;   3200 AA.
XX
AC   XXXXXX
XX
DT
XX
DE   Sequence 8, Application US/08477451
XX
CC    Sequence 8, Application US/08477451
CC    Patent No. 5928865
CC     GENERAL INFORMATION:
CC       APPLICANT:  Covacci, Antonello
CC       TITLE OF INVENTION:  Helicobacter Pylori Cagi Region
CC       NUMBER OF SEQUENCES:  46
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Chiron Corporation
CC         STREET:  4560 Horton Street
CC         CITY:  Emeryville
CC         STATE:  CA
CC         COUNTRY:  USA
CC         ZIP:  94608-2916
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/477,451
CC         FILING DATE:  07-JUN-1995
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  McClung, Barbara G.
CC         REGISTRATION NUMBER:  33,113
CC         REFERENCE/DOCKET NUMBER:  0335.002
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  510-601-2708
CC         TELEFAX:  510-655-3542
CC     INFORMATION FOR SEQ ID NO:  8:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  3200 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   3200 AA; 37997 MW; 5281128 CN;

  Query Match             29.1%;  Score 71;  DB 2;  Length
  Best Local Similarity  47.1%;  Pred. No. 7.18e+01;
  Matches    8;  Conservative    5;  Mismatches   4;

Db    401 FELLIASVFFGHYKVFL 417
          : |:|:| :|||  :|
Qy     17 YLLVISSTLFGHYIALL 33


RESULT   14
ID   US-08-469-649-2      STANDARD;      PRT;   332 AA.
XX
AC   XXXXXX
XX
DT
XX
DE   Sequence 2, Application US/08469649
XX
CC    Sequence 2, Application US/08469649
CC    Patent No. 5728562
CC     GENERAL INFORMATION:
CC       APPLICANT:  Shinyo, Tatsuo
CC       APPLICANT:  Sugihara, Keiji
CC       APPLICANT:  Takamoto, Yuji
CC       APPLICANT:  Takashio, Masachika
CC       APPLICANT:  Kamimura, Minoru
CC       APPLICANT:  Yamamoto, Kazumi
CC       APPLICANT:  Kojima, Yoshio
CC       APPLICANT:  Kikuchi, Yoshiro
CC       APPLICANT:  Emi, Shideyuki
CC       TITLE OF INVENTION:  AN ISOLATED UBICASE
CC       NUMBER OF SEQUENCES:  2
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Fitzhaut, Baldi, Goodman, Lateer
CC         ADDRESSEE:  Chick, P.C.
CC         STREET:  767 Third Avenue
CC         CITY:  New York
CC         STATE:  New York
CC         COUNTRY:  U.S.A.
CC         ZIP:  10017-2023
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  5.25 inch, 360 Kb Storage
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC DOS/MS DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/469,649
CC         FILING DATE:  June 6, 1995
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Barth, Richard
CC         REGISTRATION NUMBER:  29,180
CC         REFERENCE/DOCKET NUMBER:  950780IPD
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (212) 319-4900
CC         TELEFAX:  (212) 319-5101
CC         TELEX:  236268
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  332 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   332 AA; 37295 MW; 7809417 CN;

  Query Match             28.7%;  Score 70;  DB 1;  Length
  Best Local Similarity  40.0%;  Pred. No. 8.48e+01;
  Matches   12;  Conservative    5;  Mismatches  11;

Db    159 DNFLNITEQQSGLAGL:::KVS:NSFV:FI 188
```

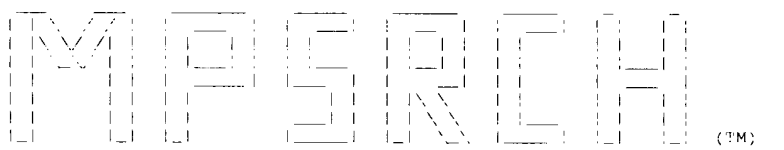
```
          | :|||  ||  |:|  |: :|:  |  :|
Qy      2 DASLNIILQQVDLIGYLLV-ISSILFGHYI 30


RESULT   15
ID   US-08-951 648-2     STANDARD;     PRT;   766 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application US/08951648
XX
CC    Sequence 2, Application US/08951648
CC    Patent No. 5932465
CC     GENERAL INFORMATION:
CC       APPLICANT:  Loughney, Kate
CC       TITLE OF INVENTION:  Phosphodiesterase 8
CC       NUMBER OF SEQUENCES:  38
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
CC         STREET:  233 South Wacker, Sears Tower Suite 6400
CC         CITY:  Chicago
CC         STATE:  Illinois
CC         COUNTRY:  US
CC         ZIP:  60606
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/951,648
CC         FILING DATE:
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Williams Jr., Joseph A.
CC         REGISTRATION NUMBER:  38,659
CC         REFERENCE/DOCKET NUMBER:  27866/34038
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  312-474-6300
CC         TELEFAX:  312-474-0448
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  766 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
CC       FEATURE:
CC         OTHER INFORMATION:
CC         OTHER INFORMATION:  /note= "The amino acid at position 290 is either Pro
CC         OTHER INFORMATION:  or Leu."
SQ   SEQUENCE   766 AA; 86842 MW; 3017162 CN;

  Query Match           28.7%;  Score 70;  DB 2;  Length 766;
  Best Local Similarity  46.4%;  Pred. No. 8.48e+01;
  Matches     8;  Conservative    6;  Mismatches   8;  Indels   0;  Gaps   0;

Db     233 QTELNDFLLDVSKTYFDNIVAI 254
           1 :|  :|| :| | | : :|:
Qy      11 QVDLIGYLLVISSILFGHYIAL 32


Search completed: Fri Aug 25 18:03:02 2000
Job time : 7 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 25 17:59:43 2000; MasPar time 3.14 Seconds
263.919 Million cell updates/sec

Tabular output not generated.

Title: >US-09-257-585-6
Description: (1-35) from US09257585.pep
Perfect Score: 244
Sequence: 1 MDASLNITLQQVDLIGYLLVISSTLFGHYIALLAA 35

Scoring table: PAM 150
Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq36
1:geneseqp

Statistics: Mean 23.550, Variance 109.579, scale 0.215

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	92	37.7	521	1	R79933	Jojoba wax synthase.	4.23e+00
2	92	37.7	521	1	R77165	Jojoba fatty acyl-CoA	4.23e+00
3	92	37.7	524	1	R79932	Jojoba wax synthase.	4.23e+00
4	92	37.7	524	1	R77164	Jojoba fatty acyl CoA	4.23e+00
5	92	37.7	524	1	R37486	Jojoba fatty acyl redu	4.23e+00
6	85	34.8	445	1	W41166	Metal-regulated transp	1.41e+01
7	83	34.0	165	1	R07360	Rat melanin-concentrat	1.97e+01
8	80	32.8	498	1	R77171	Condensing enzyme clon	3.27e+01
9	76	31.1	208	1	R77170	Lunaria condensing enz	6.36e+01
10	76	31.1	262	1	R42502	A. rhizogenes NIAES172	6.36e+01
11	76	31.1	279	1	R42506	Protein encoded by Roo	6.36e+01
12	75	30.7	250	1	R19675	Factor X LAC1 hybrid p	7.50e+01
13	75	30.7	436	1	R22513	Truncated precursor of	7.50e+01
14	75	30.7	437	1	R88468	Human two chain factor	7.50e+01
15	75	30.7	461	1	P93714	Hybrid protein of prot	7.50e+01
16	75	30.7	488	1	W76219	Human Factor X protein	7.50e+01
17	75	30.7	488	1	R22512	Mutated precursor of h	7.50e+01
18	75	30.7	488	1	W76217	Human Factor X protein	7.50e+01
19	75	30.7	488	1	W76218	Human Factor X protein	7.50e+01
20	75	30.7	488	1	W76216	Human Factor X protein	7.50e+01
21	75	30.7	488	1	R22511	Human Factor Xai.	7.50e+01
22	74	30.3	243	1	Y11116	S. pneumoniae nopaline	8.83e+01
23	72	29.5	446	1	P34292	Marine tyrosinase.	1.22e+02
24	72	29.5	446	1	P67526	Marine Mel5	1.22e+02
25	72	29.5	446	1	W72483	S. colwellii	1.22e+02
26	72	29.5	1636	1	Y00110	Enterococcus	1.22e+02
27	72	29.5	1638	1	Y03114	Enterococcus	1.22e+02
28	72	29.5	1648	1	Y00148	Enterococcus	1.22e+02
29	71	29.1	178	1	R30235	Human serin	1.44e+02
30	71	29.1	697	1	W98248	H. pylori G	1.44e+02
31	70	28.7	342	1	P33187	Amino acid	1.90e+02
32	70	28.7	942	1	Y00217	Enterococcus	1.90e+02
33	70	28.7	962	1	Y03239	Enterococcus	1.90e+02
34	68	27.9	191	1	R05306	Human calc	2.50e+02
35	68	27.9	299	1	W74841	Human secre	2.50e+02
36	68	27.9	592	1	W42453	Aquifex al	2.50e+02
37	68	27.9	671	1	W38555	Streptococ	2.50e+02
38	67	27.5	336	1	R48729	G-protein	3.30e+02
39	67	27.5	336	1	W 27761	G-protein	3.30e+02
40	67	27.5	465	1	R41973	Human neiss	3.30e+02
41	67	27.5	509	1	W96240	Human herpe	3.30e+02
42	66	27.0	82	1	W20559	H. pylori s	4.35e+02
43	66	27.0	267	1	W21027	H. pylori p	4.35e+02
44	66	27.0	315	1	W20813	H. pylori s	4.35e+02
45	66	27.0	525	1	R79466	Biosynthet	4.35e+02

ALIGNMENTS

RESULT 1
ID R79933 standard; Protein; 521 AA.
AC R79933;
DT 16-JAN-1996 (first entry)
DE Jojoba wax-synthase.
KW Wax-synthase; long-chain alcohol fatty acyltransf
KW jojoba; oilseed; transgenic plant; crop improvemen
OS Simmondsia chinensis.
PN US5445947-A.
PD 29-AUG-1995.
PF 20-NOV-1991; 796256.
PR 20-NOV-1991; US-796256.
PR 21-AUG-1992; US-933411.
PR 13-NOV-1992; WO-US09864.
PR 20-MAY-1993; US-065299.
PA (CALJ) CALGENE INC.
PI Lardizabal KD, Lassner MW, Metz JG;
DR WPI; 95-310894/40.
DR N-PSDB; I04124.
PT DNA construct expressing jojoba wax synthase and
PT Brassica cells - useful for producing wax ester(s),
PT pharmaceuticals and cosmetics, etc
PS Claim 1; Column 49-54; 50pp; English.
CC 2 Wax-synthases were isolated from a jojoba micros
CC and tryptic fragments were sequenced (R79935 42).
CC information was used to design oligonucleotide pro
CC jojoba embryo cDNA libraries. A 1.5 kb clone was
CC additional sequences were identified by 5'RACE or
CC Clone pCGN7614 was obtd. (I04124) that encoded wax
CC was used to construct plasmids for transformation
SQ Sequence 521 AA;

Query Match 37.7%; Score 92; DB 1; Leng
Best Local Similarity 44.8%; Pred. No. 4.23e+0
Matches 13; Conservative 0; Mismatches 6;

Db 32 SNLHHVKLGYHYLIGNALKVFLLLG 59
Qy 6 NITLQQVDLIGYLLVISSTLFGHYIALLAA 35

RESULT 2
ID R77165 standard; Protein; 521 AA.
AC R77165;
DT 04-DEC-1995 (first entry)
DE Jojoba fatty acyl CoA cytoplasmic protein encoded
KW Very long chain fatty acids; fatty acyl-CoA.

```
OS   Jojoba
PN   WO9515387-A.
PD   08-JUN-1995.
PF   30-NOV-1994; U13686
PR   30-NOV 1993; US-160602.
PR   23-JUN-1994; US-265047.
PA   (CALJ ) CALGENE INC.
PI   Lardizabal KD,  Lassner MW,  Metz JG,
DR   WPI; 95-215267/28.
DR   N-PSDB; Q90208.
PT   Production of very long chain fatty acid(s) in plant(s) - to produce
PT   drought and stress resistant transgenic plant(s)
PS   Example; Fig 3; 149pp; English.
CC   RNA is isolated from jojoba embryos and used to construct a cDNA
CC   library. Synthetic oligos are synthesised which corresp. to the
CC   AA sequences of selected peptide fragments and segments of DNA
CC   are produced. The DNA fragments obtd. by PCR are labeled and used
CC   as a probe to screen clones from the cDNA libraries. An approx.
CC   1500 nt. jojoba cDNA clone is obtd. in this manner. Comparison
CC   to the peptide fragments in R77175-R77182 reveals the presence of
CC   each of these peptides in the translated sequence, with the
CC   exception of SQ1129. The mRNA is approx. 2kb. Preliminary DNA
CC   sequence of a jojoba gene is presented in Q90107/R77164. Further
CC   DNA sequence analysis of additional clones indicates that there are
CC   at least 2 classes of cDNAs encoding this jojoba protein. The
CC   plasmid contg. the entire coding region in pCGN1703 is constructed
CC   to contain a SalI site approx. 8 nts 5' to the ATG start codon and
CC   is designated pCGN7614. The complete DNA sequence of pCGN7614 is
CC   given in Q90208/R77165. The major difference between the two
CC   classes of cDNAs is the presence (Q90207) or absence (Q90208) of
CC   the 6 nt coding sequence for AAs 23 and 24 of R77164.
SQ   Sequence   521 AA;

  Query Match           37.7%;  Score 92;  DB 1;  Length 521;
  Best Local Similarity  44.8%;  Pred. No. 4.23e+00;
  Matches   13;  Conservative    9;  Mismatches   6;  Indels   1;  Gaps   1;

Db      32 SINLHHVKL-GYHYLISNALFLVFIPLLG 59
           :| |::| | ||  :||::||  :|:||:
Qy       6 NIILQQVDLTGYLLVISSTLFGHYIALLA 34


RESULT    3
ID   R79932 standard; Protein; 524 AA.
AC   R79932;
DT   16-JAN-1996  (first entry)
DE   Jojoba wax-synthase.
KW   Wax-synthase; long-chain alcohol-fatty-acyl-transferase; wax ester;
KW   jojoba; oilseed; transgenic plant; crop improvement; Brassica.
OS   Simmondsia chinensis.
FH   Key             Location/Qualifiers
FT   misc_difference 213
FT                   /note= "unidentified amino acid"
PN   US5445947-A.
PD   29-AUG-1995
PF   20-NOV-1991; 796256.
PR   20-NOV-1991; US-796256.
PR   21-AUG-1992; US-933411.
PR   13-NOV-1992; WO-U09863.
PR   20-MAY-1993; US-066299.
PA   (CALJ ) CALGENE INC.
PI   Lardizabal KD,  Lassner MW,  Metz JG;
DR   WPI; 95-310894/40.
DR   N-PSDB; T04123.
PT   DNA construct expressing jojoba wax synthase and transformed
PT   Brassica cells - useful for producing wax ester(s) for use in
PT   pharmaceuticals and cosmetics, etc
PS   Claim 1; Column 47-50; 50pp; English
CC   2 Wax-synthases were isolated from a jojoba microsomal membrane prepn.
CC   and tryptic fragments were sequenced (R79932-42). The sequence
CC   information was used to design oligonucleotide probes for screening
CC   jojoba embryo cDNA libraries.  A 1.5 kb clone was isolated and
CC   additional sequences were identified by 5'RACE or by rescreening.  A
CC   1733 bp clone was obtd. (T04123) that encoded wax-synthase.
SQ   Sequence   524 AA;

  Query Match           37.7%;  Score 92;  DB 1;  Length 524;
  Best Local Similarity  44.8%;  Pred. No. 4.23e+00;
  Matches   13;  Conservative    9;  Mismatches   6;  Indels   1;  Gaps   1;

Db      34 SINLHHVKL-GYHYLISNALFLVFIPLLG 61
           :| |::| | ||  :||::||  :|:||:
Qy       6 NIILQQVDLTGYLLVISSTLFGHYIALLA 34


RESULT    4
ID   R77164 standard; Protein; 524 AA.
AC   R77164;
DT   04-DEC-1995  (first entry)
DE   Jojoba fatty acyl-CoA cytoplasmic protein.
KW   Very long chain fatty acid; fatty acyl-CoA.
OS   Jojoba.
FH   Key             Location/Qualifiers
FT   misc_difference 213
FT                   /note= "X= translation of codon 'ANC'"
PN   WO9515387-A.
PD   08-JUN-1995.
PF   30-NOV-1994; U13686.
PR   30-NOV-1993; US-160602.
PR   23-JUN-1994; US-265047.
PA   (CALJ ) CALGENE INC.
PI   Lardizabal KD,  Lassner MW,  Metz JG;
DR   WPI; 95-215267/28.
DR   N-PSDB; Q90207.
PT   Production of very long chain fatty acid(s) in plant(s) - to produce
PT   drought and stress resistant transgenic plant(s)
PS   Example; Fig 2; 149pp; English.
CC   RNA is isolated from jojoba embryos and used to construct a cDNA
CC   library. Synthetic oligos are synthesised which corresp. to the
CC   AA sequences of selected peptide fragments and segments of DNA
CC   are produced. The DNA fragments obtd. by PCR are labeled and used
CC   as a probe to screen clones from the cDNA libraries. An approx.
CC   1500 nt. jojoba cDNA clone is obtd. in this manner. Comparison
CC   to the peptide fragments in R77175-R77182 reveals the presence of
CC   each of these peptides in the translated sequence, with the
CC   exception of SQ1129. The mRNA is approx. 2kb. Preliminary DNA
CC   sequence of a jojoba gene is presented in Q90107/R77164. Further
CC   DNA sequence analysis of additional clones indicates that there are
CC   at least 2 classes of cDNAs encoding this jojoba protein. The
CC   plasmid contg. the entire coding region in pCGN1703 is constructed
CC   to contain a SalI site approx. 8 nts 5' to the ATG start codon and
CC   is designated pCGN7614. The complete DNA sequence of pCGN7614 is
CC   given in Q90208/R77165. The major difference between the two
CC   classes of cDNAs is the presence (Q90207) or absence (Q90208) of
CC   the 6 nt coding sequence for AAs 23 and 24 of R77164.
SQ   Sequence   524 AA;

  Query Match           37.7%;  Score 92;  DB 1;  Length 524;
  Best Local Similarity  44.8%;  Pred. No. 4.23e+00;
  Matches   13;  Conservative    9;  Mismatches   6;  Indels   1;  Gaps   1;

Db      34 SINLHHVKL-GYHYLISNALFLVFIPLLG 61
           :| |::| | ||  :||::||  :|:||:
Qy       6 NIILQQVDLTGYLLVISSTLFGHYIALLA 34


RESULT    5
ID   R37486 standard; Protein; 524 AA.
AC   R37486;
DT   20-SEP-1993  (first entry)
DE   Jojoba fatty acyl reductase enzyme.
KW   Tryptic peptide; crop plant; fatty acyl:fatty alcohol acyltransferase;
KW   fatty acyl reductase; jojoba; wax ester; wax synthase; cosmetics;
KW   detergents; plastics; lubricants; pharmaceuticals; sperm whale.
OS   Simmondsia chinensis.
FH   Key             Location/Qualifiers
```

FT misc_difference 213
FT /label= Ile, Thr, Asn, Ser
PN WO9310241-A.
PD 27-MAY-1993.
PF 13-NOV-1992; U09853.
PR 20-NOV-1991; US-796256.
PR 21-AUG-1992; US-933411.
PA (CALJ) CALGENE INC.
PI Lardizabal KD, Lassner MW, Metz JG;
DR WPI; 93-182556/22.
DR N-PSDB; Q42839.
PT Recombinant DNA construct for forming transgenic host cells which
PT produce wax ester(s) comprises nucleic acid sequence which
PT encodes part of wax synthase protein and heterologous DNA
PT sequence not naturally associated with wax synthase protein
PT encoding sequence
PS Disclosure; Fig 2; 79pp; English.
CC The sequences given in R47485-86 represent the jojoba wax synthase
CC (fatty acyl.fatty alcohol acyltransferase) and fatty acyl reductase
CC enzymes respectively. The DNA encoding these proteins were used in
CC a construct in which the wax synthase gene is associated with a
CC heterologous DNA sequence not naturally associated with it. This
CC construct may be used to produce crop plants having a convenient
CC source of wax esters. Wax esters can be used in a variety of
CC industrial applications, including pharmaceuticals, cosmetics,
CC detergents, plastics and lubricants. Production of wax esters in
CC crop plants allows easier isolation than from traditional sources,
CC eg. jojoba and sperm whale.
SQ Sequence 524 AA;

Query Match 37.7%; Score 92; DB 1; Length 524;
Best Local Similarity 44.8%; Pred. No. 4.2e-00;
Matches 13; Conservative 9; Mismatches 6; Indels 1; Gaps 1;

Db 34 SINLHHVKL-GYHYLISNALFLVFIPLLG 61
:| |:|| | || :||:|| :|:||:
Qy 6 NIILQQVDLTGYLLVISSILFSHYIALLA 34

RESULT 6
ID W41166 standard; Protein; 345 AA.
AC W41166;
DT 30-JUL-1998 (first entry)
DE Metal-regulated transporter polypeptide IRT2.
KW MRT; transgenic plant; pollution; metal-deficiency.
OS Arabidopsis thaliana.
PN WO9745000-A1.
PD 04-DEC-1997.
PF 27-NOV-1996; U19065.
PR 11-OCT-1996; CA-187728.
PR 29-MAY-1996; US-018578.
PA (DART-) DARTMOUTH COLLEGE.
PA (MINE) UNIV MINNESOTA.
PI Eide DJ, Guerinot ML;
DR WPI; 98-032230/03.
DR N-PSDB; V12767.
PT Novel A. thaliana metal-regulated transporter proteins and related
PT DNA - for generating transgenic plants useful for removing
PT pollutants from soil or providing nutrient to plant, suffering
PT from metal-deficiency disorders
PS Claim 33; Fig 23; 144pp; English.
CC The metal-regulated transporter (MRT) polypeptides (W41162-W41166) are
CC capable of transporting metals such as Fe(II), Pb and Zn. The peptides
CC can be used to generate transgenic plants in which the expression of MRT
CC can be altered. These plants can be used to remove pollutants from the
CC soil or to treat a disorder associated with metal-deficiency. The
CC polypeptides can also be used to identify agents which inhibit or
CC stimulate MRT or MRT expression.
SQ Sequence 345 AA;

Query Match 34.8%; Score 85; DB 1; Length 345;
Best Local Similarity 46.7%; Pred. No. 1.4le-01;
Matches 11; Conservative 10; Mismatches 9; Indels 0; Gaps 0;

Db 40 PLKIVAIVAILIISLIGVLSFRSRYISFL 69
:|:|: || |: :|| |:|:|:|:|
Qy 4 SLNIILQQVDLTGYLLVISSILFSHYIALL 33

RESULT 7
ID R07360 standard; protein; 165 AA.
AC R07360;
DT 29-JAN-1991 (first entry)
DE Rat melanin-concentrating precursor.
KW Melanin concentrating hormone; skin disorders; me
KW ACTH secretion.
FH Key Location/Qual ifiers
FT region 141. .149
FT /label= MCH peptide
PN WO9011295-A.
PD 04-OCT-1990.
PF 20-MAR-1990; U01492.
PR 22-MAR-1989; US-326984.
PA (SALK) SALK INST FOR BIOL STUD.
PI Vaughan J, Fischer WH, Rivier JE, Nahon JM, Presse
DR WPI; 90-320225/42.
DR N-PSDB; Q06245.
PT Cyclic mammalian hormone for concentrating mammali
PT comprises peptide based on 19 amino acid residues
PT linkages.
PS Disclosure; page 34; 47pp; English.
CC This is therat melanin concentrating hormone (MCH)
CC See also Q06238-42, Q06244 and Q06246-48.
SQ Sequence 165 AA;

Query Match 34.8%; Score 85; DB 1; Leng
Best Local Similarity 57.9%; Pred. No. 1.97e-0
Matches 10; Conservative 1; Mismatches 8;

Db 1 MAKMSLSSYMLMLAFSLFSHCI ..SA 26
: : |::|:::: :| :| : ::
Qy 9 LQQVDLTGYLLVISSILFSHYIALLAA 35

RESULT 8
ID R77171 standard; Protein; 4[illegible] AA.
AC R77171;
DT 04-DEC-1995 (first entry)
DE Condensing enzyme clone Lunaria ..
KW Lunaria; condensing enzyme.
OS Lunaria annua.
FH Key Location/Qualifiers
FT misc_difference 102
FT /note= "[illegible]ation of codon. N"
PN WO9515387-A.
PD 08-JUN-1995.
PF 30-NOV-1994; U13686.
PR 30-NOV-1993; US-160062.
PR 23-JUN-1994; US-265047.
PA (CALJ) CALGENE INC.
PI Lardizabal KD, Lassner MW, Metz JG;
DR WPI; 95-212267/28.
DR N-PSDB; Q90217.
PT Production of very long chain fatty acid(s) in pla
PT drought and stress resistant transgenic plant(s)
PS Example; Figure 12; 149pp; English.
CC The CE15 and CE20 Brassica cDNA sequences (See Q90
CC and Q90212) and the condensing enzyme encoding seq
CC jojoba (Q90208) were used in determining primers
CC Q90225 from conserved AAs. These primers were vari
CC PCR (RTPCR) amplify fragments from mRNA isolated fr
CC seeds of Lunaria annua, Tropaeolum majus (Nasturtiu
CC green liliques of Arabidopsis thaliana. The primer
CC successfully utilised were Q90221 and Q90222. Thes
CC used to produce three clones encoding a portion of
CC condensing enzyme from Arabidopsis, From Lunaria a

```
CC   was identified, LUN CE8. A cDNA library from RNA isolated from
CC   developing seeds of Lunaria was constructed, and LUN CE8 was used
CC   to screen this library. Three classes of cDNA clones were isolated,
CC   Lunaria 1, Lunaria 5 and Lunaria 27. Lunaria 5 shares approx.
CC   85% homology with the Brassica CE20 clones, and it is suggested
CC   that it is active in seed oil fatty acid elongation.
SQ   Sequence   498 AA;

  Query Match           32.8%;  Score 80;  DB 1;  Length 498;
  Best Local Similarity  29.0%;  Pred. No. 3.27e-01;
  Matches    9;  Conservative   14;  Mismatches    7;  Indels    1;  Gaps    1;

Db      56 LDLSLQLQFNLVGFIFFITASIFGFTVIFMS 86
           |:: || | :| |::: |:::|| : :::
Qy       5 LNIILQ-QVDLIGYLLVISSILFGHYIALLA 34


RESULT    9
ID   R77170 standard; Protein; 208 AA.
AC   R77170;
DT   04-DEC-1995  (first entry)
DE   Lunaria condensing enzyme clone LUN CE8.
KW   lunaria; condensing enzyme.
OS   Lunaria annua.
PN   WO9515387-A.
PD   08-JUN-1995.
PF   30-NOV-1994; 013686.
PR   30-NOV-1993; US-160602.
PR   23-JUN-1994; US-265947.
PA   (CALJ ) CALGENE INC.
PI   Lardizabal KD,  Lassner MW,  Metz JG;
DR   WPI; 95-215267/28.
DR   N-PSDB; Q90216.
PT   Production of very long chain fatty acid(s) in plant(s)   to produce
PT   drought and stress resistant transgenic plant(s)
PS   Example; Figure 11; 149pp; English.
CC   The CE15 and CE20 Brassica cDNA sequences (see Q90210,Q90211
CC   and Q90212) and the condencing enzyme encoding sequence from
CC   jojoba (Q90208) were used in determining primers Q90221-
CC   Q90225 from conserved AAs. These primers were variously used to
CC   PCR (RTPCR) amplify fragments from RNA isolated from developing
CC   seeds of Lunaria annua, Tropaeolu majus (Nasturtium), and
CC   green liliques of Arabidopsis thaliana. The primers most
CC   successfully utilised were Q90221 and Q90222. These primers were
CC   used to produce three clones encoding a portion of the elongase
CC   condensing enzyme from Arabidopsis. From Lunaria a single clone
CC   was identified, LUN CE8. A cDNA library from RNA isolated from
CC   developing seeds of Lunaria was constructed, and LUN CE8 was used
CC   to screen this library. Three classes of cDNA clones were isolated,
CC   Lunaria 1, Lunaria 5 and Lunaria 27. Lunaria 5 shares approx.
CC   85% homology with the Brassica CE20 clones, and it is suggested
CC   that it is active in seed oil fatty acid elongation.
SQ   Sequence   208 AA;

  Query Match           31.1%;  Score 76;  DB 1;  Length 208;
  Best Local Similarity  29.0%;  Pred. No. 6.36e+01;
  Matches    9;  Conservative   13;  Mismatches    8;  Indels    1;  Gaps    1;

Db      39 LDLSLQLQFNLVGFIFFITVSIFGFTVIFMS 69
           |:: || | :| |::: |: ::|| : :::
Qy       5 LNIILQ-QVDLTGYLLVISSTLFGHYIALLA 34


RESULT   10
ID   R42502 standard; Protein; 262 AA.
AC   R42502;
DT   22-FEB-1994  (first entry)
DE   A. rhizogenes NIAES1724 megalo-plasmid coded protein.
KW   Dwarfing; rooting; plants; breeding; disease prevention;
KW   identification; classification; pathogens.
OS   Agrobacterium rhizogenes NIAES1724.
PN   JO5192156-A.
PD   03-AUG-1993.
PF   10 DEC 1991, 360827.
PR   10-DEC-1991; JP-360827.
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (SUMQ ) SUMITOMO METAL IND LTD.
DR   WPI; 93-277471/35.
DR   N PSDB; Q46099.
PT   DNA having specified base sequence relating to dwarfing and
PT   rooting of plants - is used for breeding and prevention of diseases
PT   of plants e.g. identification and classification of pathogens
PS   Example, Page 9-11; 19pp; Japanese.
CC   The sequence is that of a protein encoded by DNA from an Agrobacterium
CC   rhizogenes NIAES1724 megalo-plasmid relating to the dwarfing and
CC   rooting of plants.
SQ   Sequence   262 AA;

  Query Match           31.1%;  Score 76;  DB 1;  Length 262;
  Best Local Similarity  37.5%;  Pred. No. 6.36e+01;
  Matches    9;  Conservative    9;  Mismatches    6;  Indels    0;  Gaps    0;

Db     155 VEEPDLEGFIAIFASTLFIHLLEL 178
           ::: || |:: :::|||| | : |
Qy       9 LQQVDLTGYLLVISSTLFGHYIAL 32


RESULT   11
ID   R42506 standard; Protein; 279 AA.
AC   R42506;
DT   16-FEB-1994  (first entry)
DE   Protein encoded by Root-induction gene rolB from A.rhizogenes.
KW   hairy root induction; Ri plasmid; T-DNA; plant cell.
OS   Agrobacterium rhizogenes MAFF03-01724.
PN   JO5192154-A.
PD   03-AUG-1993.
PF   28 JAN 1992, 013058.
PR   29 JAN-1991; JP-009385.
PA   (DAIL ) DAICEL CHEM IND LTD.
DR   WPI; 93-277469/35.
DR   N PSDB; Q46113.
PT   Hair like root inducing gene present on T-DNA region of
PT   Ri plasmid   for artificially controlling induction of hair type
PT   root in plant cell
PS   Claim 3; Page 2; 17pp; Japanese.
CC   The rolB gene was localised to a 951bp NaeI SplI restriction
CC   fragment of A.rhizogenes I DNA. Three rol ("root induction") genes
CC   were isolated which can be used to control hairy root induction in
CC   plant cells. See also Q46112 and Q46114.
SQ   Sequence   279 AA;

  Query Match           31.1%;  Score 76;  DB 1;  Length 279;
  Best Local Similarity  37.5%;  Pred. No. 6.36e+01;
  Matches    9;  Conservative    9;  Mismatches    6;  Indels    0;  Gaps    0;

Db     172 VEEPDLEGFIAIFASTLFIHLLEL 195
           ::: || |:: :::|||| | : |
Qy       9 LQQVDLTGYLLVISSTLFGHYIAL 32


RESULT   12
ID   R13675 standard; Protein; 250 AA.
AC   R13675;
DT   09-OCT-1991  (first entry)
DE   Factor X-LACI hybrid protein.
KW   Kunitz domain; blood; coagulation; inhibitor; Factor X;
KW   tissue factor; TF; Lipoprotein-Associated Coagulation Inhibitor.
FH   Key             Location/Qualifiers
FT   protein         1. .171
FT                   /label= X1c
FT   peptide         1. .40
FT                   /label= prepro_leader
FT   domain          55. .64
FT                   /label= GLA_domain
FT   domain          89. .150
FT                   /label= growth_factor_domains
```

```
FT   domain            172. .250
FT                     /label= kunitz_domain
FT   disulfide_bond    57. .62
FT   disulfide_bond    90. .101
FT   disulfide_bond    95. .110
FT   disulfide_bond    112. .121
FT   disulfide_bond    129. .149
FT   disulfide_bond    146. .149
FT   disulfide_bond    151. .164
FT   disulfide_bond    186. .236
FT   disulfide_bond    195. .219
FT   disulfide_bond    211. .232
PN   EP-439442-A.
PD   31-JUL-1991.
PF   21-JAN-1991; 870008.
PR   25-JAN-1990; US-470289.
PA   (UNIW ) UNIV OF WASHINGTON.
PI   Girard TJ, Broze GJ;
DR   WPI; 91-224839/31.
DR   N-PSDB; Q12776.
PT   New factor X-LACI hybrid protein - comprises light chain of
PT   factor X and LACI's first kunitz domain for use as anticoagulant
PS   Disclosure; Page 12-14; 17pp; English.
CC   The protein is used as a blood coagulation inhibitor in mammals.  It
CC   is believed to mimic the Xa/LACI complex in binding to and
CC   inhibiting VIIa/tissue factor.  LACI inhibits via a novel feedback
CC   mechanism requiring generation of Xa (a prod. of VIIa/TF activity);
CC   XlcLACIKi inhibits VIIa/TF activity directly
CC   The DNA allows prodn. of XlcLACIki by introduction of the gene into
CC   cells suitable for expression, e.g. E. coli or CHO cells.
SQ   Sequence   250 AA;

  Query Match             30.7%;  Score 75;  DB 1;  Length 250;
  Best Local Similarity   34.6%;  Pred. No. 7.50e+01;
  Matches    9;  Conservative    9;  Mismatches   7;  Indels   1;  Gaps   1;

Db       1 MGRPLHLVLLSASLAG LLLLGESLF 25
           1  :|::::|    |:| :|::: :|:
Qy       1 MDASLNIILQQVDLTGYLLVISSTLF 26


RESULT   13
ID   R22513 standard; Protein; 436 AA.
AC   R22513;
DT   28-JUL-1992  (first entry)
DE   Truncated precursor of human Factor Xai.
KW   Mutant; prothrombinase complex; proteolytic; precursor; thrombosis;
KW   inflammation; restenosis; transplantation; haemophilia; antibodies.
OS   Homo sapiens.
PN   WO9204378-A.
PD   19-MAR-1992.
PF   04-SEP-1991; 006347.
PR   04-SEP-1990; US-578646.
PA   (CORT-) COR THERAPEUTICS IN.
PI   Wolf D;
DR   WPI; 92-114303/14.
PT   New analoques of Factor XA peptide - useful for treating
PT   haemophilia, thrombosis, inflammation and transplant
PT   complications, for in-vivo diagnosis
PS   Claim 7; Fig 1; 59pp; English.
CC   The full length cDNA of human factor X (Mpl9X) was converted to
CC   encode a truncated form of human Factor X, designated rX', by deletion
CC   of the activation peptide by oligonucleotide site directed mutagenesis.
CC   An oligonucleotide was used to align Arg 142 following the
CC   C-terminus of the Factor X light chain with Ile 53 of the Factor X
CC   activation peptide (1st residue of the heavy chain). When expressed
CC   in CHO cells the truncated peptide was cleaved endogenously.
CC   Modified Factor Xa was further produced by acylation e.g. with the
CC   p-nitrophenyl ester of p-toluoylic acid.  Factor rX' is used to
CC   treat or prevent thrombosis; inflammation; restenosis or complications
CC   of transplantation.  It is also used in treatment of adult respiratory
CC   distress syndrome and haemophilia.  The modified Factor X has no
CC   proteolytic activity and interferes with the ability of endogenous
CC   factor Xa to convert pro-thr into to thrombin.  An
CC   with Factor rX' are passive therapeutic agents and
CC   See also R22512.
SQ   Sequence   436 AA;

  Query Match             30.7%;  Score 75;  DB 1;  Length
  Best Local Similarity   34.6%;  Pred. No. 7.50e+01;
  Matches    9;  Conservative    9;  Mismatches

Db       1 MGRPLHLVLLSASLAG LLLLGESLF 25
           1  :|::::|    |:| :|::: :|:
Qy       1 MDASLNIILQQVDLTGYLLVISSTLF 26


RESULT   14
ID   R88468 standard; protein; 4-  AA.
AC   R88468;
DT   30-AUG-1996  (first entry)
DE   Human two chain factor Xai single chain precursor
KW   Human; factor X; single chain; factor X'i; precurs
KW   factor Xai; hexapeptide linker; Asn88; Ala185; pro
KW   no enzymatic activity; haemostasis regulation; p
KW   treatment; thrombus formation; vasculature pathol
KW   thrombin induced; restenosis; inflammation; procoa
KW   wound healing; bypass factor; replacement therapy;
KW   imaging; diagnosis; antibody production.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   peptide          1. .40
FT                    /note= "prepro leader sequence"
FT   protein          41. .177
FT                    /note= "light chain"
FT   peptide          178. .184
FT                    /note= "hexapeptide linker"
FT   protein          184. .437
FT                    /note= "heavy chain"
FT   domain           57. .62
FT                    /note= "GLA domain"
FT   domain           90. .184
FT                    /note= "growth factor domains"
FT   misc_difference  101
FT                    /note= "given as beta in specific
FT   misc_difference  271
FT                    /note= "wild type Asp substd. with
FT   misc_difference  368
FT                    /note= "wild type Ser substd. with
FT   disulfide_bond   56. .66
FT   disulfide_bond   88. .99
FT   disulfide_bond   93. .104
FT   disulfide_bond   110. .119
FT   disulfide_bond   127. .138
FT   disulfide_bond   134. .147
FT   disulfide_bond   149. .162
FT   disulfide_bond   170. .291
FT   disulfide_bond   190. .195
FT   disulfide_bond   210. .226
FT   disulfide_bond   339. .364
FT   disulfide_bond   364. .392
PN   WO9600577-A1.
PD   11-JAN-1996.
PF   28-JUN-1995; 008368.
PR   29-JUN-1994; US-268008.
PA   (CORT-) COR THERAPEUTICS INC.
PI   Sinha U,  Wolf DL;
DR   WPI; 96-077335/08.
PT   New pro-coagulant composition to treat haemophilia
PT   modified blood factors selected from Factor Xa, Fa
PT   VIIa and activated Protein C
PS   Example 3; Fig 3; 69pp; English.
CC   The present protein is the human factor X derived,
CC   factor X'i, a precursor for the two chain factor X.
CC   differs from factor X in that the factor X activat
CC   been replaced with a hexapeptide linker, and the a
```

```
CC   Asp88 and Ser185 have been substd. with Asn and Ala, respectively.
CC   The hexapeptide linker comprises a proteolytic cleavage site, which
CC   is cleaved to activate factor X'i, thus becoming the two chain
CC   factor Xai. Factor Xai unlike its activated factor X wild type
CC   homologue, factor Xa, has no enzymatic activity due to the
CC   proteolytic domain amino acid substitutions Asn88 and Ala185.
CC   Factor Xai can be used for the regulation of haematostasis and for
CC   the prevention and treatment of thrombus formation and other
CC   pathological processes in the vasculature induced by thrombin, e.g.
CC   restenosis and inflammation. It may be used as a procoagulant, e.g.
CC   for wound healing, as a bypass factor, in replacement therapy,
CC   haemophilia treatment, in imaging and other diagnostic uses and in
CC   the prodn. of antibodies which can be used in diagnosis and
CC   therapy.
SQ   Sequence   437 AA;

  Query Match             30.7%;  Score 75;  DB 1;  Length 437;
  Best Local Similarity   34.6%;  Pred. No. 7.50e+01;
  Matches     9;  Conservative    9;  Mismatches   7;  Indels   1;  Gaps   1;

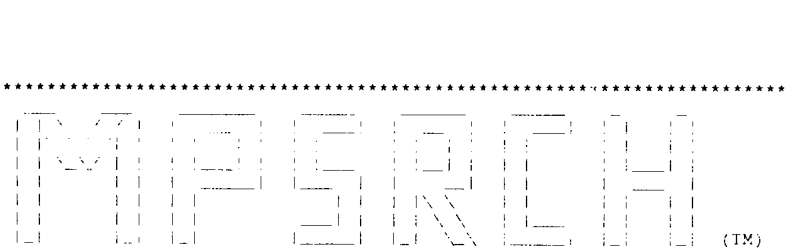
Db       1 MGRPLHLVLLSASLAG-LLLLGESLF 25
           |  :|:::|    |:| ||::: :||
Qy       1 MDASLNIILQQVDLTGYLLVISSILF 26


RESULT   15
ID   P93714 standard; protein; 461 AA.
AC   P93714;
DT   4-JUN-1990  (first entry)
DE   Hybrid protein of protein-C and Factor-X.
KW   Fusion protein; anticoagulant; protein-C; Factor-X; Gla domain; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   peptide         1. .40
FT                   /label=signal_peptide
PN   EP-296413-A.
PD   28-DEC-1988.
PF   09-SEP-1988; 109186.
PR   12-JUN-1987; JP-145293.
PA   (FARH) Hoechst Japan Ltd.
PI   Iwasaki W, Takahashi M, Hashimoto T;
DR   WPI; 89-000910/01.
DR   N-PSDB; N94063.
PT   Hybrid protein of protein-C with replaced Gla domain - using human
PT   vitamin K dependent proteins, eg Factor-X, to give improved
PT   anticoagulation activity.
PS   Disclosure;  Page 16-19; 23pp; English.
CC   The sequence is that of a fusion protein in which the Gla domain of
CC   protein is replaced with that of Factor X. The novel protein has a more
CC   potent anticoagulation activity than protein-C and is expected to have
CC   competitive inhibitory activity against menadione dependent blood
CC   coagulation proteins. It also has neutralization activity against
CC   plasminogen activation inhibitor, and inactivates Factor-Va or
CC   Factor-VIIIa.
SQ   Sequence 461 AA;

  Query Match             30.7%;  Score 75;  DB 1;  Length 461;
  Best Local Similarity   34.6%;  Pred. No. 7.50e+01;
  Matches     9;  Conservative    9;  Mismatches   7;  Indels   1;  Gaps   1;

Db       1 MGRPLHLVLLSASLAG-LLLLGESLF 25
           |  :|:::|    |:| ||::: :||
Qy       1 MDASLNIILQQVDLTGYLLVISSILF 26


Search completed: Fri Aug 25 18:00:04 2000
Job time : 21 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 25 17:57:38 2000; MasPar time 9.40 Seconds
501.644 Million cell updates/sec

Tabular output not generated.

Title: >US-09-257-585-5
Description: (1-68) from US09257585.pep
Perfect Score: 522
Sequence: 1 MAGLINEDFIFLNSYLFVPI.....RKGDKGGECSLDFHCGLFQL 68

Scoring table: PAM 150
Gap 11

Searched: 225878 seqs, 69331122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl12
1:sp_archea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 38.589; Variance 80.103; scale 0.482

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	94	18.0	361	7	Q31066	MHC PD1 MAJOR TRANSPLA	5.39e-01
2	94	18.0	361	7	O19244	MHC CLASS I ANTIGEN.	5.39e-01
3	94	18.0	363	7	Q31067	MHC PD1A MAJOR TRANSPL	5.39e-01
4	94	18.0	364	7	O19245	MHC CLASS I ANTIGEN.	5.39e-01
5	94	18.0	364	7	Q31068	MHC PD14 TRANSPLANTATI	5.39e-01
6	94	18.0	366	7	Q31069	MHC PD14A TRANSPLANTAT	5.39e-01
7	93	17.8	361	7	O19242	MHC CLASS I ANTIGEN.	7.27e-01
8	92	17.6	437	8	Q34191	NADH DEHYDROGENASE SUB	9.79e-01
9	92	17.6	509	8	Q9XMR6	NADH DEHYDROGENASE SUB	9.79e-01
10	91	17.4	167	5	Q20714	CODED FOR BY C. ELEGAN	1.32e+00
11	91	17.4	301	5	Q22568	F19C4.3 PROTEIN.	1.32e+00
12	91	17.4	364	7	O19243	MHC CLASS I ANTIGEN.	1.32e+00
13	91	17.4	527	2	O25926	ALGINATE O ACETYLATION	1.32e+00
14	91	17.4	537	1	Q9YAP8	537AA LONG HYPOTHETICA	1.32e+00
15	91	17.4	1360	5	O02096	FRUIT FLY RETROTRANSPO	1.32e+00
16	90	17.2	283	2	O51521	CONSERVED HYPOTHETICAL	1.77e+00
17	90	17.2	317	5	P92015	HYPOTHETICAL 35.7 KD P	1.77e+00
18	90	17.2	364	7	O19240	MHC CLASS I ANTIGEN.	1.77e+00
19	89	17.0	611	5	O97264	PFC0625W PROTEIN.	2.32e+00
20	88	16.9	473	1	O59352	473AA LONG HYPOTHETICA	3.16e+00
21	87	16.7	93	2	Q43439	MNXF.	4.14e+00
22	87	16.7	106	2	Q9X774	HYPOTHETICAL	4.14e+00
23	87	16.7	521	1	Q24543	PUTATIVE AMI	4.14e+00
24	86	16.5	91	7	[illegible]	MHC CLASS I	5.41e+00
25	86	16.5	293	2	Q58014	F55C9.4 PR	5.41e+00
26	86	16.5	744	2	[illegible]	[illegible]	5.41e+00
27	85	16.3	488	10	Q22952	HYPOTHETICAL	7.07e+00
28	85	16.3	500	1	Q47514	HYPOTHETICAL	7.07e+00
29	85	16.3	512	12	Q55437	TERPENE CY	7.07e+00
30	84	16.1	165	2	Q58589	NADH DEHYI	9.23e+00
31	84	16.1	195	8	Q7Y588	NADH DEHYI	9.23e+00
32	84	16.1	264	5	Q21883	[illegible]5	9.23e+00
33	84	16.1	294	5	Q94295	C05H11.1	9.23e+00
34	84	16.1	442	5	Q17888	F49E5.6	9.23e+00
35	84	16.1	461	7	O14241	MHC CLASS	9.23e+00
36	84	16.1	487	8	Q34575	NADH DEH	9.23e+00
37	84	16.1	671	2	Q42E42	HYPOTHET	9.23e+00
38	83	15.9	182	1	O51523	HYPOTHET	1.20e+01
39	83	15.9	428	1	P80066	42DAA I	1.20e+01
40	83	15.9	433	2	Q54147	S.MILAB	1.20e+01
41	83	15.9	440	5	Q21719	C05MIL	1.20e+01
42	83	15.9	494	2	Q59080	LANOSTE	1.20e+01
43	83	15.9	667	8	Q21277	HAEM 1	1.20e+01
44	83	15.9	1041	10	Q48947	[illegible]	1.20e+01
45	82	15.7	424	2	Q9WX66	SIMILA	1.56e+01

ALIGNMENTS

RESULT 1
ID Q31066 PRELIMINARY; PRT; 361 AA
AC Q31066;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last Sequence
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotati
DE MHC PD1 MAJOR TRANSPLANTATION ANTIGEN.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; V
OC Eutheria; Cetartiodactyla; Suina; Suidae;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 85282928.
RA SATZ M.L., WANG L.C., SINGER D.S., MARK
RT "Structure and expression of two porcine
RT I MHC antigens.";
RL J. Immunol. 134:2167-2175(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 88174731.
RA EHRLICH K., MACQUIRE J.E., SINGER D.S.;
RT "Identification of negative and positi
RT associated with a class I major histo
RL Mol. Cell. Biol. 8:695-703(1988).
DR EMBL; M21057; AAA30076.1; -.
DR HSSP; P13747; 1A9E.
DR PROSITE; PS00290; IG_MHC; 1.
DR PFAM; PF00047; ig; 1.
DR PFAM; PF00129; MHC_I; 1.
KW MHC.
SQ SEQUENCE 361 AA; 40244 MW; [illegible]

Query Match 18.0%; Score [illegible]
Best Local Similarity 34.2%; Pred. N
Matches [illegible]; Conservative [illegible]; M

Db [illegible]

RESULT 2
ID [illegible] PRELIMINARY; PRT
AC [illegible];

```
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   MHC CLASS I ANTIGEN.
GN   PD1.
OS   Sus scrofa (Piq).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   SULLIVAN J.A., OETTINGER H.F., SACHS D.H., EDGE A.S.B.;
RL   J. Immunol. 159:0-0(1997).
DR   EMBL; AF014005; AAB69340.1; -.
DR   HSSP; P30685; 1A9E.
DR   PROSITE; PS00290; IG_MHC; 1.
DR   PFAM; PF00047; iq; 1.
DR   PFAM; PF00129; MHC_I; 1.
KW   MHC.
SQ   SEQUENCE   361 AA;  40214 MW;  2F952C07 CRC32;

  Query Match           18.0%;  Score 94;  DB 7;  Length 361;
  Best Local Similarity  34.2%;  Pred. No. 5.39e-01;
  Matches   13;  Conservative   13;  Mismatches  10;  Indels   2;  Gaps   2;

Db     302 VPIVGI-IVGL-VLVLVAGAMVAGVVIWRKTRSGEKGG 337
           ||| :| :| : |:: :  |: :::  |: :| |:||| 
Qy      18 VPIGSIYVVDIAVPITLDDAIFPSIGCWKVSRKGDKGG 55


RESULT    3
ID   Q31067      PRELIMINARY;      PRT;   363 AA.
AC   Q31067;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   MHC PD1A MAJOR TRANSPLANIATION ANTIGEN.
OS   Sus scrofa (Piq).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 85262925.
RA   SATZ M.L., WANG L.C., SINGER D.S., RUDIKOFF S.,
RT   "Structure and expression of two porcine genomic clones encoding class
RT   I MHC antigens.";
RL   J. Immunol. 135:2167-2175(1985).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88174711.
RA   EHRLICH R., MAGUIRE J.E., SINGER D.S.;
RT   "Identification of negative and positive regulatory elements
RT   associated with a class I major histocompatibility complex gene.";
RL   Mol. Cell. Biol. 8:695-703(1988).
DR   EMBL; M21057; AAA31077.1; -.
DR   HSSP; P30685; 1A9E.
DR   PROSITE; PS00290; IG_MHC; 1.
DR   PFAM; PF00047; iq; 1.
DR   PFAM; PF00129; MHC_I; 1.
KW   MHC.
SQ   SEQUENCE   363 AA;  40400 MW;  FBD87172 CRC32;

  Query Match           18.0%;  Score 94;  DB 7;  Length 363;
  Best Local Similarity  34.2%;  Pred. No. 5.39e-01;
  Matches   13;  Conservative   14;  Mismatches  10;  Indels   2;  Gaps   2;

Db     302 VPIVGI-IVGL-VLVLVAGAMVAGVVIWRKTRSGEKGG 337
           ||| :| :| : |:: :  |: :::  |: :| |:||| 
Qy      18 VPIGSIYVVDIAVPITLDDAIFPSIGCWKVSRKGDKGG 55


RESULT    4
ID   O19245      PRELIMINARY;      PRT;   364 AA.
AC   O19245;
```

```
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN 1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   MHC CLASS I ANTIGEN.
GN   PD14.
OS   Sus scrofa (Piq).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   SULLIVAN J.A., OETTINGER H.F., SACHS D.H., EDGE A.S.B.;
RL   J. Immunol. 159:0-0(1997).
DR   EMBL; AF014006; AAB69341.1; -.
DR   HSSP; P30460; 1AGE.
DR   PROSITE; PS00290; IG_MHC; 1.
DR   PFAM; PF00047; iq; 1.
DR   PFAM; PF00129; MHC_I; 1.
KW   MHC.
SQ   SEQUENCE   364 AA;  40507 MW;  06D14FB3 CRC32;

  Query Match           18.0%;  Score 94;  DB 7;  Length 364;
  Best Local Similarity  34.2%;  Pred. No. 5.39e-01;
  Matches   13;  Conservative   12;  Mismatches  11;  Indels   2;  Gaps   2;

Db     305 VPIVGI-IVGL-VLVLVAGAVVAGVVIWRKKRSGEKGG 340
           ||| :| :| : |:: :  |: :::  |:  | |:||| 
Qy      18 VPIGSIYVVDIAVPITLDDAIFPSIGCWKVSRKGDKGG 55


RESULT    5
ID   Q31068      PRELIMINARY;      PRT;   364 AA.
AC   Q31068;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   MHC PD14 TRANSPLANTATION ANTIGEN.
OS   Sus scrofa (Piq).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 85262925.
RA   SATZ M.L., WANG L.C., SINGER D.S., RUDIKOFF S.;
RT   "Structure and expression of two porcine genomic clones encoding class
RT   I MHC antigens.";
RL   J. Immunol. 135:2167-2175(1985).
DR   EMBL; M21058; AAA31078.1; -.
DR   HSSP; P30491; 1A1O.
DR   PROSITE; PS00290; IG_MHC; 1.
DR   PFAM; PF00047; ig; 1.
DR   PFAM; PF00129; MHC_I; 1.
KW   MHC.
SQ   SEQUENCE   364 AA;  40605 MW;  D6FEF016 CRC32;

  Query Match           18.0%;  Score 94;  DB 7;  Length 364;
  Best Local Similarity  34.2%;  Pred. No. 5.39e-01;
  Matches   13;  Conservative   12;  Mismatches  11;  Indels   2;  Gaps   2;

Db     305 VPIVGI-IVGL-VLVLVAGAVVAGVVIWRKKRSGEKGG 340
           ||| :| :| : |:: :  |: :::  |:  | |:||| 
Qy      18 VPIGSIYVVDIAVPITLDDAIFPSIGCWKVSRKGDKGG 55


RESULT    6
ID   Q31069      PRELIMINARY;      PRT;   366 AA.
AC   Q31069;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   MHC PD14A TRANSPLANTATION ANTIGEN.
OS   Sus scrofa (Piq).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 85262925.
RA   SATZ M.I., WANG [illegible] SINGER D.S., RUDIKOFF S.;
RT   "Structure and expression of two porcine genomic clones encoding class
RT   I MHC antigens.";
RL   J. Immunol. 135:2167-2175(1985).
DR   EMBL; M21058; AAA31079.1; -.
DR   HSSP; P30491; 1ALO.
DR   PROSITE; PS00290; IG_MHC; 1.
DR   PFAM; PF00047; ig; 1.
DR   PFAM; PF00129; MHC_I; 1.
KW   MHC.
SQ   SEQUENCE   366 AA;  40796 MW;  CE91BE21 CRC32;

  Query Match           18.0%;  Score 94;  DB 7;  Length 366;
  Best Local Similarity  34.2%;  Pred. No. 5.39e-01;
  Matches   14;  Conservative   12;  Mismatches  11;  Indels   2;  Gaps   2;

Db    305 VPIVGI-IVGL-VLVLVAGAVVAGVVIWRKKRSGEKGG 340
          ||| :| :| : ||: :  |: :::  |:  | |:|||
Qy     18 VPIGSIYVVDIAVFITLDDAIFPSIGCWKVSRKGDKGG 55


RESULT    7
ID   O19242      PRELIMINARY;      PRT;   361 AA.
AC   O19242;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   MHC CLASS I ANTIGEN.
GN   PC1.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   SULLIVAN J.A., OETTINGER H.F., SACHS D.H., EDGE A.S.B.;
RL   J. Immunol. 159:0-0(1997).
DR   EMBL; AF014003; AAB69338.1; -.
DR   HSSP; P30460; 1AGE.
DR   PROSITE; PS00290; IG_MHC; 1.
DR   PFAM; PF00047; ig; 1.
DR   PFAM; PF00129; MHC_I; 1.
KW   MHC.
SQ   SEQUENCE   361 AA;  40234 MW;  030E9556 CRC32;

  Query Match           17.8%;  Score 93;  DB 7;  Length 361;
  Best Local Similarity  34.2%;  Pred. No. 7.27e-01;
  Matches   14;  Conservative   12;  Mismatches  11;  Indels   2;  Gaps   2;

Db    302 VPIVGI-IVGL-VLVLVAGAVVAGVVIWRKKRSGEKGG 337
          ||| :| :| : ||: :  |: :::  |:  | |:|||
Qy     18 VPIGSIYVVDIAVFITLDDAIFPSIGCWKVSRKGDKGG 55


RESULT    8
ID   Q34191      PRELIMINARY;      PRT;   437 AA.
AC   Q34191;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   NADH DEHYDROGENASE SUBUNIT 4.
GN   ND4.
OS   Crithidia oncopelti.
OG   Mitochondrion.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Crithidia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S-068;
RA   DMITRIY MASLOV A.;
RL   Submitted (OCT-1990) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S-068;
RA   MASLOV D.A., HORVATH A., WANG [illegible] K., KOLESNIKOV V.
RL   Submitted (APR-1993) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X56015; CAA39491.1; -.
KW   Mitochondrion.
SQ   SEQUENCE   437 AA;  [illegible] MW;  741F2793 CRC32;

  Query Match           17.6%;  Score 92;  DB 8;  Len
  Best Local Similarity  34.4%;  Pred. No. 9.79e-01;
  Matches   11;  Conservative   10;  Mismatches  [illegible]

Db    365 GLISLSFIYIICFYVIVLSVFVSSIYIYICL 395
          |||:  |||:  : :| : |::|  | ::| |
Qy      3 GLINEDFIFLNSYLFVPIGSIYVVDIAVFITL 34


RESULT    9
ID   Q9XMR6      PRELIMINARY;      PRT;   509 AA.
AC   Q9XMR6;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   NADH DEHYDROGENASE SUBUNIT 4 (EC 1.6.5.3).
GN   NAD4.
OS   Tetrahymena pyriformis.
OG   Mitochondrion.
OC   Eukaryota; Alveolata; [illegible]; Oligohymenophorea
OC   Tetrahymenina; Tetrahymena.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   EDQVIST J., BURGER G., GRAY M.W.;
RT   "Expression of mitochondrial protein coding genes
RT   pyriformis.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ dat
RN   [2]
RP   SEQUENCE FROM N.A.
RA   BURGER G., ZHU Y., [illegible]EN I.G., GREENWOOD S.J
RA   LANG B.F., GRAY M.W.;
RT   "Complete sequence, gene content and organization
RT   genome of Tetrahymena pyriformis. Comparison with
RT   mitochondrial DNA.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ dat
DR   EMBL; AF160864; AAD41955.1; -.
KW   Oxidoreductase; Mitochondri
SQ   SEQUENCE   509 AA;  [illegible] MW;  7EE9294C CRC32;

  Query Match           17.6%;  Score 92;  DB 8;  Len
  Best Local Similarity  36.6%;  Pred. No. 9.79e-01;
  Matches   11;  Conservative   15;  Mismatches   8;

Db    283 IYLSGFLVKTALTGFFKLNLSFIVVSSIFIAICL 317
          |:|:::|  :: :::  :: :: |  :  ::| :| |
Qy     10 IFLNSYLF VPIGSIYVVDIAVFITLDAIFPSIGC 44


RESULT   10
ID   Q20714      PRELIMINARY;      PRT;   162 AA.
AC   Q20714;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   CODED FOR BY C. ELEGANS cDNA YK61C2.5.
GN   F53B3.5.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdi
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C.,
RA   BONFIELD J., BURTON J., CONNELL M., COUSEY T., Cr
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A.
```

```
RA   CARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN J., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   MILLER N.;
RL   Submitted (NOV 1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   WATERSTON R.;
RL   Submitted (NOV 1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U40414; AAA81409.1; -.
SQ   SEQUENCE   167 AA;  18813 MW;  A19FDA47 CRC32;

  Query Match           17.4%;  Score 91;  DB 5;  Length 167;
  Best Local Similarity  51.9%;  Pred. No. 1.32e+00;
  Matches   14;  Conservative    4;  Mismatches   8;  Indels   2;  Gaps   2;

Db      39 YVVGI-VFVVLS-FIVGLIGCWNRSRK 63
           ||| | ||: |  :  |  |||| : ||
Qy      24 YVVDIAVFITLDDAIFPSIGCWKVSRK 50


RESULT   11
ID   Q22568      PRELIMINARY;      PRT;   301 AA.
AC   Q22568;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   T19C4.3 PROTEIN.
GN   T19C4.3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   WHITE S.;
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSEN N.F., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN J., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z75549; CAA99915.1; -.
SQ   SEQUENCE   301 AA;  34861 MW;  96293B92 CRC32;

  Query Match           17.4%;  Score 91;  DB 5;  Length 301;
  Best Local Similarity  29.0%;  Pred. No. 1.32e+00;
  Matches    9;  Conservative   15;  Mismatches   5;  Indels   2;  Gaps   2;

Db     172 IPDTFIYLIPFMVVVSVSSV-LVNVISFVSL 201
           | : ||:| ::: | :::|: :|::: |::|
Qy       5 INFDFIFLNSYLFV-PIGSIYVVDIAVFITL 34
```

```
RESULT   12
ID   Q19243      PRELIMINARY;      PRT;   364 AA.
AC   Q19243;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   MHC CLASS I ANTIGEN.
GN   PC14.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   SULLIVAN J.A., GETTINGER H.F., SACHS D.H., EDGE A.S.B.;
RL   J. Immunol. 159:0-0(1997).
DR   EMBL; AF014004; AAB69339.1; -.
DR   HSSP; P10685; 1A9E.
DR   PROSITE; PS00290; IG_MHC; 1.
DR   PFAM; PF00047; ig; 1.
DR   PFAM; PF00129; MHC_I; 1.
KW   MHC.
SQ   SEQUENCE   364 AA;  40791 MW;  BFC101BB CRC32;

  Query Match           17.4%;  Score 91;  DB 7;  Length 364;
  Best Local Similarity  34.2%;  Pred. No. 1.32e+00;
  Matches   13;  Conservative   12;  Mismatches   11;  Indels   2;  Gaps   2;

Db     305 VPIVGI-IVGL-VLVLVAGAMVAGVVIWRKKRSGEKGG 340
           ||| :| :| : |:: :  |: ::: |:  | |:|||
Qy      18 VPIGSIYVVDIAVFITLDDAIFPSIGCWKVSRKGDKGG 55


RESULT   13
ID   O25526      PRELIMINARY;      PRT;   527 AA.
AC   O25526;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   ALGINATE O ACETYLATION PROTEIN (ALGI).
GN   HP0855.
OS   Helicobacter pylori (Campylobacter pylori).
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=26695;
RX   MEDLINE; 97394467.
RA   TOMB J.-F., WHITE O., KERLAVAGE A.R., CLAYTON R.A., SUTTON G.G.,
RA   FLEISCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,
RA   NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,
RA   LOFTUS B., RICHARDSON D., DODSON R., KHALAK H.G., GLODEK A.,
RA   MCKENNEY K., FITZEGERALD L.M., LEE N., ADAMS M.D., HICKEY E.K.,
RA   BERG D.E., GOCAYNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,
RA   COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATTHEY L., WALLIN E.,
RA   HAYES W.S., BOROD0VSKY M., KARP P.D., SMITH H.O., FRASER C.M.,
RA   VENTER J.C.;
RT   "The complete genome sequence of the gastric pathogen Helicobacter
RT   pylori.";
RL   Nature 388:539-547(1997).
DR   EMBL; AE000596; AAD07902.1; -.
DR   TIGR; HP0855; -.
KW   Hypothetical protein.
SQ   SEQUENCE   527 AA;  61344 MW;  8D272D66 CRC32;

  Query Match           17.4%;  Score 91;  DB 2;  Length 527;
  Best Local Similarity  34.3%;  Pred. No. 1.32e+00;
  Matches   12;  Conservative    9;  Mismatches   13;  Indels   1;  Gaps   1;

Db     329 FLKEYLYIPLGGNRVKELIVYRNLI LVFLIGGFW 362
           ||: ||:::|:|:  | :: |:  |   :|   | |
Qy      11 FLNSYLFVPIGSIYVVDIAVFITLDDAIFPSIGCW 45
```

```
RESULT   14
ID   Q9YAP8       PRELIMINARY;      PRT:   537 AA.
AC   Q9YAP8;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   537AA LONG HYPOTHETICAL PROTEIN.
GN   APE1895.
OS   Aeropyrum pernix.
OC   Archaea; Crenarchaeota; Aeropyrum.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K1;
RX   MEDLINE; 99310339.
RA   KAWARABAYASI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HARIKAWA Y.,
RA   JIN-NO K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,
RA   HOSOYAMA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,
RA   TAKAMIYA M., MASUDA S., FUNAHASHI T., TANAKA T., KUDOH Y.,
RA   YAMAZAKI J., KUSHIDA N., OGUCHI A., AOKI K., KUBOTA K., NAKAMURA Y.,
RA   NOMURA N., SAKO Y., KIKUCHI H.;
RT   "Complete genome sequence of an aerobic hyper-thermophilic
RT   crenarchaeon, Aeropyrum pernix K1.";
RL   DNA Res. 6:83-101(1999).
DR   EMBL; AP000062; BAA80900.1; -.
SQ   SEQUENCE   537 AA;  58871 MW;  23AD0086 CRC32;

  Query Match             17.4%;  Score 91;  DB 1;  Length 537;
  Best Local Similarity   40.0%;  Pred. No. 1.32e-00;
  Matches    10;  Conservative    9;  Mismatches   5;  Indels   1;  Gaps   1;

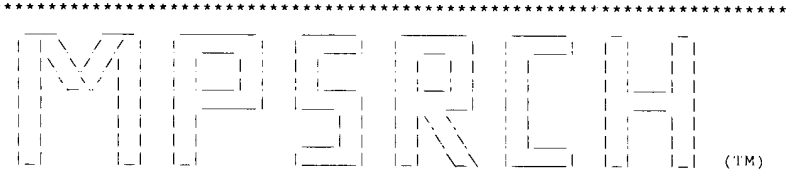
Db    499 YFYLALYLLLSAGSIFVAVEIIDFV 523
          :::|  ||::: |||:| |:|  |:
Qy      9 FIFLNSYLFVPIGSIYV-VDIAVFI 32


RESULT   15
ID   O02006       PRELIMINARY;      PRT:  1360 AA.
AC   O02006;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   FRUIT FLY RETROTRANSPOSON NINJA DNA, COMPLETE SEQUENCE (FRAGMENT).
OS   Drosophila simulans (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WHITE-CHOCOLATE; TISSUE=WHOLE BODY; TRANSPOSON=NINJA;
RX   MEDLINE; 96252483.
RA   OGURA K., TAKECHI S., NAKAYAMA I., YAMAMOTO M.;
RT   "Molecular structure of the transposable element ninja in Drosophila
RT   simulans.";
RL   Genes Genet. Syst. 71:1-8(1996).
DR   EMBL; D83207; BAA19771.1; -.
DR   FLYBASE; FBgn0015168; Dsim\ninja.
DR   PFAM; PF00098; zf-CCHC; 1.
FT   NON_TER       1      1
SQ   SEQUENCE   1360 AA;  150935 MW;  CC5DE16F CRC32;

  Query Match             17.4%;  Score 91;  DB 5;  Length 1360;
  Best Local Similarity   28.6%;  Pred. No. 1.32e-00;
  Matches     8;  Conservative   12;  Mismatches   8;  Indels   0;  Gaps   0;

Db   1275 GAVRAVELHVFVDASQAAFAAVAYWRVI 1302
          ::  |:: ||:   :| |:::: |:|:
Qy     21 GSIYVVDIAVFITLDDAIPPSIGCWKVS 48


Search completed: Fri Aug 25 17:58:48 2000
Job time : 70 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 25 17:57:02 2000; MasPar time 4.21 Seconds
500.730 Million cell updates/sec

Tabular output not generated

Title: >US-09-257-585-5
Description: (1-68) from US09257585.pep
Perfect Score: 522
Sequence: 1 MAGLINEDFIFLNSYLEVPI.....KAGDAGGPCSLDFHCGLFQL 68

Scoring table: PAM 150
Gap 11

Searched: 85661 seqs, 30989116 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 39.542; Variance 76.101; scale 0.520

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	99	19.0	873	1	VGLB_ILTV6	GLYCOPROTEIN B PRECURS	2.81e-02
2	99	19.0	883	1	VGLB_ILTVT	GLYCOPROTEIN B PRECURS	2.81e-02
3	94	18.0	112	1	OL7D_MOUSE	OLFACTORY RECEPTOR 7D	1.41e-01
4	94	18.0	883	1	VGLB_ILTVS	GLYCOPROTEIN B PRECURS	1.41e-01
5	93	17.8	138	1	YIGG_ECOLI	HYPOTHETICAL 15.8 KDA	1.94e-01
6	93	17.8	344	1	IN37_SPIOL	37 KDA INNER ENVELOPE	1.94e-01
7	93	17.8	366	1	MRAY_RHIMP	PHOSPHO-N-ACETYLMURAMO	1.94e-01
8	91	17.4	411	1	SECY_PORPU	PREPROTEIN TRANSLOCASE	3.61e-01
9	90	17.2	347	1	OPSD_NECAU	RHODOPSIN (FRAGMENT).	4.96e-01
10	89	17.0	366	1	1C02_GORGO	CLASS I HISTOCOMPATIBI	6.75e-01
11	89	17.0	366	1	1C03_GORGO	CLASS I HISTOCOMPATIBI	6.75e-01
12	89	17.0	366	1	1C04_GORGO	CLASS I HISTOCOMPATIBI	6.75e-01
13	89	17.0	809	1	NAH2_RABIT	SODIUM/HYDROGEN EXCHAN	6.75e-01
14	89	17.0	813	1	NAH2_RAT	SODIUM/HYDROGEN EXCHAN	6.75e-01
15	88	16.9	425	1	SYS_BORBU	SERYL-TRNA SYNTHETASE	9.17e-01
16	88	16.9	832	1	TRNL_CANAL	TRNA LIGASE (EC 6.5.1.	9.17e-01
17	88	16.9	877	1	WEE1_SCHPO	MITOSIS INHIBITOR PROT	9.17e-01
18	87	16.7	182	1	CD3G_HUMAN	T-CELL SURFACE GLYCOPR	1.24e+00
19	87	16.7	365	1	1C01_GORGO	CLASS I HISTOCOMPATIBI	1.24e+00
20	87	16.7	503	1	CP38_MACFA	CYTOCHROME P450 3A8 (E	1.24e+00
21	87	16.7	901	1	SOK1_YEAST	SOK1 PROTEIN.	1.24e+00
22	86	16.5	377	1	Y147_MYCPN	HYPOTHETICAL PROTEIN M	1.68e+00
23	85	16.3	101	1	NU5M_LEITA	NADH-UBIQUINONE OXIDOR	2.27e+00
24	85	16.3	176	1	NU6C_ORYSA	NADH-PLASTO	2.27e+00
25	85	16.3	1683	1	YJD9_YEAST	HYPOTHETICA	2.27e+00
26	84	16.1	199	1	YAF7_CAEEL	HYPOTHETICA	3.07e+00
27	84	16.1	721	1	NU1M_NEUCR	NADH-UBIQUI	3.07e+00
28	84	16.1	418	1	NTBX_CAITY	PUTATIVE	3.07e+00
29	84	16.1	1675	1	[illegible]_LAST	1,3 BETA-GL	3.07e+00
30	83	15.9	242	1	ATP6_RICPR	ATP SYNTHA	4.16e+00
31	83	15.9	312	1	PUR3_VIGUN	PHOSPHORIB	4.16e+00
32	83	15.9	1042	1	VG37_BPI4	BASEPLATE S	4.16e+00
33	83	15.9	1102	1	YK04_CAEEL	HYPOTHETICA	4.16e+00
34	83	15.9	1895	1	GLS2_YEAST	1,3 BETA-GL	4.16e+00
35	82	15.7	228	1	Y072_METJA	HYPOTHETICA	5.64e+00
36	82	15.7	499	1	YWBE_BACSU	HYPOTHETICA	5.64e+00
37	81	15.5	210	1	ALKH_IREPA	PUTATIVE KE	7.37e+00
38	81	15.5	38	1	[illegible]_CAEEL	RHODOPSIN (	7.37e+00
39	81	15.5	422	1	TERL_BPSPP	TERMINASE L	7.37e+00
40	81	15.5	426	1	YHFN_BACSU	HYPOTHETICA	7.37e+00
41	81	15.5	465	1	ALSI_BACSU	AMINO ACID	7.37e+00
42	81	15.5	556	1	NU2M_PODAN	NADH-UBIQUI	7.37e+00
43	81	15.5	652	1	NU5M_PODAN	NADH-UBIQUI	7.37e+00
44	81	15.5	716	1	NU5M_NEUCR	NADH-UBIQUI	7.37e+00
45	80	15.3	342	1	DPXX_HUMAN	HLA CLASS I	9.63e+00

ALIGNMENTS

RESULT 1
ID VGLB_ILTV6 STANDARD; PRT; 873 AA.
AC Q02409;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE GLYCOPROTEIN B PRECURSOR.
GN GB.
OS Infectious laryngotracheitis virus (strain 632) (
OC Viruses; dsDNA viruses, no RNA stage; Herpesvirida
OC Alphaherpesvirinae; Varicellovirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92180317.
RA Poulsen D., Burton C.R., O'Brian J.J., Jeffrey J.
RA Keeler J., Calvin L.;
RT "Identification of the infectious laryngotracheit
RT glycoprotein qB gene by the polymerase chain react
RL Virus Genes 5:335-347(1992).
RN [2]
RP REVISIONS.
RA Keeler J.;
RL Submitted (FEB-1994) to the EMBL/GenBank/DDBJ data
CC -!- SIMILARITY: TO OTHER HERPESVIRUSES GLYCOPROTEI
CC ---------- [illegible]
CC This SWISS-PROT entry is copyright. It is produced [illegible]
CC between the Swiss Institute of Bioinformatics an[illegible]
CC the European Bioinformatics Institute. There are [illegible]
CC use by non-profit institutions as long as its [illegible]
CC modified and this statement is not removed. Usage [illegible]
CC entities requires a license agreement (See http://[illegible]
CC or send an email to license@isb-sib.ch).
CC ---------- [illegible]
DR EMBL; X56093; CAA39573.1; -.
DR PIR; S26690; S26690.
DR PFAM; PF00606; Glycoprotein_B; 1.
KW Signal; Glycoprotein; Transmembrane.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 873 GLYCOPROTEIN B.
FT DOMAIN 22 714 EXTRACELLULAR (BY SIM
FT TRANSMEM 715 740 BY SIMILARITY.
FT TRANSMEM 741 761 BY SIMILARITY.
FT DOMAIN 762 855 CYTOPLASMIC (BY SIMIL
FT CARBOHYD 92 92 POTENTIAL.
FT CARBOHYD 111 111 POTENTIAL.
FT CARBOHYD 201 201 POTENTIAL.
FT CARBOHYD 252 252 POTENTIAL.

```
FT   CARBOHYD     350     350       POTENTIAL.
FT   CARBOHYD     569     569       POTENTIAL.
FT   CARBOHYD     625     625       POTENTIAL.
FT   CARBOHYD     639     639       POTENTIAL.
SQ   SEQUENCE   873 AA;  99039 MW;  F73E7E6D7A0E3AAD CRC64;

  Query Match             19.0%;  Score 99;  DB 1;  Length 873;
  Best Local Similarity   24.3%;  Pred. No. 2.81e-02;
  Matches     9;  Conservative    16;  Mismatches   11;  Indels    1;  Gaps    1;

Db     607 LLGENYLLYEDYTFVRQVNASEIEEVSIFINLNATIL 643
           |: |:::: : | ||  : :  : ::::|| |: :|:
Qy       4 LINEDFIFLNSYLFV-PIGSIYVVDIAVFITLDDAIF 39


RESULT    2
ID   VGLB_ILTVT     STANDARD;      PRT;   883 AA.
AC   P24904;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   GLYCOPROTEIN B PRECURSOR.
GN   GB.
OS   Infectious laryngotracheitis virus (strain Thorne V882) (ILTV).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae, Varicellovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91132136.
RA   Griffin A.M.;
RT   "The nucleotide sequence of the glycoprotein gB gene of infectious
RT   laryngotracheitis virus: analysis and evolutionary relationship to
RT   the homologous gene from other herpesviruses.";
RL   J. Gen. Virol. 72:393-398(1991).
CC   -!- SIMILARITY: TO OTHER HERPESVIRUSES GLYCOPROTEIN B.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D00818; BAA00699.1; -.
DR   PIR; A38478; VGBEIL.
DR   PIR; JQ0951; JQ0951.
DR   PFAM; PF00606; Glycoprotein_B; 1.
KW   Signal; Glycoprotein; Transmembrane.
FT   SIGNAL        1      26       POTENTIAL.
FT   CHAIN        27     883       GLYCOPROTEIN B.
FT   DOMAIN       27     724       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    725     744       POTENTIAL.
FT   TRANSMEM    751     771       POTENTIAL.
FT   DOMAIN      772     865       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD    102     102       POTENTIAL.
FT   CARBOHYD    121     121       POTENTIAL.
FT   CARBOHYD    211     211       POTENTIAL.
FT   CARBOHYD    262     262       POTENTIAL.
FT   CARBOHYD    360     360       POTENTIAL.
FT   CARBOHYD    579     579       POTENTIAL.
FT   CARBOHYD    635     635       POTENTIAL.
FT   CARBOHYD    649     649       POTENTIAL.
SQ   SEQUENCE   883 AA;  100160 MW;  D457DF5178840A7D CRC64;

  Query Match             19.0%;  Score 99;  DB 1;  Length 883;
  Best Local Similarity   24.3%;  Pred. No. 2.81e-02;
  Matches     9;  Conservative    16;  Mismatches   11;  Indels    1;  Gaps    1;

Db     617 LLGENYLLYEDYTFVRQVNASEIEEVSIFINLNATIL 653
           |: |:::: : | ||  : :  : ::::|| |: :|:
Qy       4 LINEDFIFLNSYLFV-PIGSIYVVDIAVFITLDDAIF 39


RESULT    3
ID   OL7D_MOUSE     STANDARD;      PRT;   112 AA.
AC   Q60884;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   OLFACTORY RECEPTOR 7D (M15) (FRAGMENT).
GN   OLFR7.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria, Rodentia, Sciurognathi, Muridae, Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J;
RX   MEDLINE; 96149403.
RA   Sullivan S.L., Adamson M.C., Ressler K.J., Kozak C.A., Buck L.B.;
RT   "The chromosomal distribution of mouse odorant receptor genes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:884-888(1996).
CC   -!- FUNCTION: PUTATIVE ODORANT RECEPTOR.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- MISCELLANEOUS: OLFACTORY RECEPTORS 7A-7I ARE PRODUCED BY EIGHT
CC       DIFFERENT GENES WITHIN THE OLFR7 COMPLEX.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U28773; AAC52396.1; -.
DR   GCRDB; GCR_1688; -.
DR   MGD; MGI:104712; OLFR7.
DR   PFAM; PF00001; 7tm_1; 1.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; PARTIAL.
KW   G-protein coupled receptor; Transmembrane; Multigene family;
KW   Olfaction.
FT   NON_TER       1      1
FT   DOMAIN       <1     12       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     13     33       4 (POTENTIAL).
FT   DOMAIN       34     73       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     74     95       5 (POTENTIAL).
FT   DOMAIN       96    109       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    110   >112       6 (POTENTIAL).
FT   NON_TER     112    112
SQ   SEQUENCE   112 AA;  12429 MW;  08040529B1B851DE CRC64;

  Query Match             18.0%;  Score 94;  DB 1;  Length 112;
  Best Local Similarity   34.9%;  Pred. No. 1.41e-01;
  Matches    15;  Conservative    15;  Mismatches   12;  Indels    1;  Gaps    1;

Db      43 KSDLINHYFCDLISILNLSCSNIFVNDLVILIFSLINTIFPTL 85
           :: ||| |  | | | | :: ::|:| |: ::| :| ::|||::
Qy       1 MACLINEDFIFLNSYLFVPIGSIYVVDIAVFI TLDDAIFPSI 42


RESULT    4
ID   VGLB_ILTVS     STANDARD;      PRT;   883 AA.
AC   P27415;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   01-AUG-1992 (Rel. 23, Last annotation update)
DE   GLYCOPROTEIN B PRECURSOR.
GN   GB.
OS   Infectious laryngotracheitis virus (strain SA-2) (ILTV).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Varicellovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91335774.
```

```
RA   Kongsuwan K., Prideaux C.T., Johnson M.A., Sheppard M., Fahey K.J.;
RT   "Nucleotide sequence of the gene encoding infectious
RT   laryngotracheitis virus glycoprotein B.";
RL   Virology 184:404-410(1991).
CC   -!- SIMILARITY: TO OTHER HERPESVIRUSES GLYCOPROTEIN B.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M64927; AAA88009.1; ALT_SEQ.
DR   PIR; A40567; VGBEIS.
DR   PFAM; PF00606; Glycoprotein_B; 1.
KW   Glycoprotein; Transmembrane; Signal.
FT   SIGNAL        1     26       POTENTIAL.
FT   CHAIN        27    883       GLYCOPROTEIN B.
FT   DOMAIN       27    724       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    725    744       POTENTIAL.
FT   TRANSMEM    751    771       POTENTIAL.
FT   DOMAIN      772    883       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD    102    102       POTENTIAL.
FT   CARBOHYD    121    121       POTENTIAL.
FT   CARBOHYD    211    211       POTENTIAL.
FT   CARBOHYD    262    262       POTENTIAL.
FT   CARBOHYD    360    360       POTENTIAL.
FT   CARBOHYD    579    579       POTENTIAL.
FT   CARBOHYD    635    635       POTENTIAL.
FT   CARBOHYD    649    649       POTENTIAL.
SQ   SEQUENCE   883 AA;  100142 MW;  B4140833460550I4 CRC64;

  Query Match             18.0%;  Score 94;  DB 1;  Length 883;
  Best Local Similarity   24.3%;  Pred. No. 1.41e-01;
  Matches    9;  Conservative   15;  Mismatches  12;  Indels   1;  Gaps   1;

Db    617 LLGENYLLYEDYTFVRQVNASEIEEVSTFINLNATIL 653
          |: |:::: : | ||  : :  : ::: || |: :|:
Qy      4 LINEDFIFLNSYLFV-PIGSIYVVDIAVFITLDDAIF 39


RESULT    5
ID   YIGG_ECOLI     STANDARD;      PRT;   138 AA.
AC   P27843;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   HYPOTHETICAL 15.8 KDA PROTEIN IN COPA-PAPD INTERGENIC REGION.
GN   YIGG.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RA   Ohmori H.;
RL   Submitted (JAN 1993) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE; 92358234.
RA   Daniels D.L., Plunkett G. III, Burland V.D., Blattner F.R.;
RT   "Analysis of the Escherichia coli genome: DNA sequence of the region
RT   from 84.5 to 86.5 minutes.";
RL   Science 257:771-778(1992).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L02122; AAD15[illegible]; -.
DR   EMBL; M87049; AAA67[illegible]; -.
DR   EMBL; AE000457; AAC[illegible]; -.
DR   PIR; S30745; S30745.
DR   PIR; S30708; S30708.
DR   ECOGENE; EG11465; yigG.
KW   Hypothetical protein.
SQ   SEQUENCE   138 AA;  15794 MW;  1E7B8859EEC1E27A CR

  Query Match             17.8%;  Score 93;  DB 1;  Leng
  Best Local Similarity   28.1%;  Pred. No. 1.94e-01;
  Matches    9;  Conservative   12;  Mismatches  16;

Db     76 LRDSCFIYIKIYVFGIL-AVVIISIIIMIAED 106
          | :  |||:::|:|   : :::: |   |: |
Qy      4 LINEDFIFLNSYLFVPIGSIYVVDIAVFITLD 35


RESULT    6
ID   IN37_SPIOL     STANDARD;      PRT;   344 AA.
AC   P23525;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   37 KDA INNER ENVELOPE MEMBRANE PROTEIN, CHLOROPLAS
OS   Spinacia oleracea (Spinach).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyt
OC   Magnoliophyta; eudicotyledons; Caryophyllidae; Ca
OC   Chenopodiaceae; Spinacia.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91146572.
RA   Dreses-Werringloer U., Fischer K., Wachter E., Li
RA   Fluegge U.I.;
RT   "cDNA sequence and deduced amino acid sequence of
RT   the 37 kDa inner envelope membrane polypeptide fr
RT   chloroplasts. Its transit peptide contains an amp
RT   as the only detectable structural element.";
RL   Eur. J. Biochem. 195:461-468(1991).
RN   [2]
RP   SEQUENCE OF 120-141 AND [illegible]
RX   MEDLINE; 91348205.
RA   Block M.A., Joyard J., [illegible]
RT   "Purification and characterization of E37, a major
RT   envelope protein.";
RL   FEBS Lett. 287:167-170(1991).
CC   -!- FUNCTION: NOT YET KNOWN, ONE OF THE MAJOR PROT
CC       MEMBRANE OF THE CHLOROPLAST ENVELOPE.
CC   -!- SUBCELLULAR LOCATION: CHLOROPLAST INNER MEMBRA
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced
CC   between the Swiss Institute of Bioinformatics an
CC   the European Bioinformatics Institute. There are
CC   use by non-profit institutions as long as its
CC   modified and this statement is not removed. Usage
CC   entities requires a license agreement (See http://
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X56963; CAA40285.1; -.
DR   PIR; S14409; S14409.
DR   PFAM; PF01209; Ubie_methyltran; 1.
KW   Chloroplast; Transit peptide; Inner membrane.
FT   TRANSIT       1     21       CHLOROPLAST.
FT   CHAIN        22    344       37 KDA INNER ENVELOPE
SQ   SEQUENCE   344 AA;  38976 MW;  [illegible]

  Query Match             17.8%;  Score 93;  DB 1;  Leng
  Best Local Similarity   31.3%;  Pred. No. 1.94e-01;
  Matches   10;  Conservative    6;  Mismatches   9;
```

```
Db     308 LVFLYRFLLGALASIYYVLVPIYMWIKDKIFP 339
           ::||  :|: ::: | | ::::: : | |||
Qy       9 FIFLNSYLFVPIGSIYVVDIAVFITLDDAIFP 40


RESULT    7
ID   MRAY_RHIME     STANDARD;      PRT;   366 AA.
AC   Q52952;
DT   15-FEB-2000 (Rel. 39, Created)
DT   15-FEB-2000 (Rel. 39, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   PHOSPHO-N-ACETYLMURAMOYL-PENTAPEPTIDE-TRANSFERASE (EC 2.7.8.13)
DE   (UDP-MURNAC-PENTAPEPTIDE PHOSPHOTRANSFERASE).
GN   MRAY.
OS   Rhizobium meliloti (Sinorhizobium meliloti).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Sinorhizobium
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1021;
RA   Galibert F., Capela D., Hubler-Barloy F., Gatius M., Batut J.,
RA   Boistard P., Gouzy J., Kahn D., Thebault P., Gottesman A.,
RA   Purnelle B., Pohl T., Bothe G., Schneider S., Portetelle D.,
RA   Vandenbol M., Puehler A., Becker A., Weidner S.;
RL   Submitted (MAR-2000) to the SWISS-PROT data bank.
RN   [2]
RP   PRELIMINARY SEQUENCE OF 284-366 FROM N.A.
RC   STRAIN=1021;
RX   MEDLINE; 95011665.
RA   Leach F., Wacks D.B., Signer E.R.;
RT   "Rhizobium meliloti homologs of Escherichia coli mur genes.";
RL   Gene 148:87-90(1994).
CC   -!- FUNCTION: FIRST STEP OF THE LIPID CYCLE REACTIONS IN THE
CC       BIOSYNTHESIS OF THE CELL WALL PEPTIDOGLYCAN (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: UDP-N-ACETYLMURAMOYL-L-ALANYL-D-GLUTAMYL-L-
CC       LYSYL-D-ALANYL-D-ALANINE   UNDECAPRENYL PHOSPHATE   UMP +
CC       N ACETYLMURAMOYL L ALANYL D GLUTAMYL L LYSYL D ALANYL D ALANINE
CC       DIPHOSPHOUNDECAPRENOL.
CC   -!- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 4. MRAY
CC       SUBFAMILY.
CC   -!- CAUTION: IN ADDITIONS TO THE CONFLICT SHOWN BELOW, REF.2 SEQUENCE
CC       HAD 5 FRAMESHIFT ERRORS SCATTERED ALONG THE SEQUENCE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L25875; AAA66471.1; ALT_FRAME.
DR   PROSITE; PS01347; MRAY_1; 1.
DR   PROSITE; PS01348; MRAY_2; 1.
KW   Peptidoglycan synthesis; Cell division; Transferase; Transmembrane.
FT   TRANSMEM     27     47       POTENTIAL.
FT   TRANSMEM     71     91       POTENTIAL.
FT   TRANSMEM     93    113       POTENTIAL.
FT   TRANSMEM    138    158       POTENTIAL.
FT   TRANSMEM    174    194       POTENTIAL.
FT   TRANSMEM    205    225       POTENTIAL.
FT   TRANSMEM    245    265       POTENTIAL.
FT   TRANSMEM    268    288       POTENTIAL.
FT   TRANSMEM    297    317       POTENTIAL.
FT   TRANSMEM    343    363       POTENTIAL.
FT   CONFLICT    255    255       MISSING (IN REF. 2).
SQ   SEQUENCE   366 AA;  39265 MW;  F42FB33AAB5491CF CRC64;

  Query Match           17.8%;  Score 93;  DB 1;  Length 366;
  Best Local Similarity  35.7%;  Pred. No. 1.94e-01;
  Matches   20;  Conservative   14;  Mismatches  18;  Indels   4;  Gaps   4;

Db      74 MGGLMILTGIVVSSLLWADLSSIYVVS-TLLVTLGFGAIGFYD-DYLKVTKQSEKG 127
           |:|||:    | ::| |   ::|||||  ::::||   || |      ||:: ::||
Qy       1 MAGLINEDFIFLNSYLFVPIGSIYVVDIAVFITLD-DAI-FPSIGCWKVSRKGDKG 54


RESULT    8
ID   SECY_PORPU     STANDARD;      PRT;   411 AA.
AC   P51297;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   PREPROTEIN TRANSLOCASE SECY SUBUNIT.
GN   SECY.
OS   Porphyra purpurea.
OG   Chloroplast.
OC   Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AVONPORT;
RA   Reith M.E., Munholland J.;
RT   "Complete nucleotide sequence of the Porphyra purpurea chloroplast
RT   genome.";
RL   Plant Mol. Biol. Rep. 13:333-335(1995).
CC   -!- FUNCTION: INVOLVED IN PROTEIN EXPORT. PROBABLY INTERACTS WITH
CC       OTHER PROTEINS TO ALLOW THE TRANSLOCATION OF PROTEINS ACROSS THE
CC       CHLOROPLAST ENDOPLASMIC RETICULUM (CER) MEMBRANES.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE SECY/SEC61 ALPHA FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U38804; AAC08183.1; -.
DR   MENDEL; 10361; PoRpu;secY;1.
DR   PFAM; PF00344; secY; 1.
DR   PRINTS; PR00303; SECYTRNLCASE.
DR   PROSITE; PS00755; SECY_1; 1.
DR   PROSITE; PS00756; SECY_2; 1.
KW   Protein transport; Transmembrane; Chloroplast; Translocation.
SQ   SEQUENCE   411 AA;  45650 MW;  A353DAF0827863FD CRC64;

  Query Match           17.4%;  Score 91;  DB 1;  Length 411;
  Best Local Similarity  33.3%;  Pred. No. 3.64e-01;
  Matches   14;  Conservative   12;  Mismatches  16;  Indels   0;  Gaps   0;

Db     269 LTQLTQNTFLLQVLYLFCPNGSLYLVLYSVLILFFSYFYTSI 310
           :: |  : |::   ||| | ||:|:|  :|:| :   :: ||
Qy       1 MAGLINEDFIFLNSYLFVPIGSIYVVDIAVFITLDDAIFPSI 42


RESULT    9
ID   OPSD_NEOAU     STANDARD;      PRT;   347 AA.
AC   P79809;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   RHODOPSIN (FRAGMENT).
GN   RHO.
OS   Neoniphon aurolineatus.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Acanthopterygii;
OC   Percomorpha; Beryciformes; Holocentridae; Holocentrinae; Neoniphon.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EYE;
RA   Toller W.W., Moses K., McFall-Ngai M.J.;
RL   Submitted (JAN 1997) to the EMBL/GenBank/DDBJ databases.
```

```
CC   -!- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
CC       MEDIATE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY
CC       LINKED TO CIS-RETINAL.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: ROD SHAPED PHOTORECEPTOR CELLS WHICH MEDIATES
CC       VISION IN DIM LIGHT.
CC   -!- PTM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY
CC       BE PHOSPHORYLATED (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC       OPSIN SUBFAMILY.
CC   ---------- ---------- ----- ---------- ----------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------
DR   EMBL; U57541; AAB39522.1; -.
DR   GCRDB; GCR_1605; -.
DR   PFAM; PF00001; 7tm_1; 1.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
DR   PROSITE; PS00238; OPSIN; 1.
KW   Photoreceptor; Retinal protein; Transmembrane; Glycoprotein; Vision;
KW   Phosphorylation; Lipoprotein; Palmitate; G-protein coupled receptor.
FT   NON_TER       1      1
FT   DOMAIN       <1      33        EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      34      58        1 (POTENTIAL).
FT   DOMAIN        59      70        CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      71      95        2 (POTENTIAL).
FT   DOMAIN        96     110        EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     111     130        3 (POTENTIAL).
FT   DOMAIN       131     149        CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     150     173        4 (POTENTIAL).
FT   DOMAIN       174     199        EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     200     227        5 (POTENTIAL).
FT   DOMAIN       228     249        CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     250     273        6 (POTENTIAL).
FT   DOMAIN       274     281        EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     282     306        7 (POTENTIAL).
FT   DOMAIN       307     347        CYTOPLASMIC (POTENTIAL).
FT   DISULFID     107     184        BY SIMILARITY.
FT   BINDING      293     293        RETINAL CHROMOPHORE (BY SIMILARITY).
FT   CARBOHYD      12      12        POTENTIAL.
FT   CARBOHYD     197     197        POTENTIAL.
FT   LIPID        320     320        PALMITATE (BY SIMILARITY).
SQ   SEQUENCE   347 AA;  38991 MW;  89F6FD5FD79EA8E5 CRC64;

  Query Match             17.2%;  Score 90;  DB 1;  Length 347;
  Best Local Similarity   32.4%;  Pred. No. 4.96e-01;
  Matches    11;  Conservative   11;  Mismatches   10;  Indels    2;  Gaps    2;

Db      28 LVNPAAFAFLGAYMFFLIIIGFPVNFLTLYVILE 61
           |:|   | || :|:|  |   : |: ::::::|:
Qy       4 LIN-EDFIFLNSYLFVPIGSIYVVD-IAVFITLD 35


RESULT   10
ID   1C02_GORGO     STANDARD;      PRT:   366 AA.
AC   P30385;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   CLASS I HISTOCOMPATIBILITY ANTIGEN, GOGO-C0201 ALPHA CHAIN PRECURSOR.
OS   Gorilla gorilla gorilla (Lowland gorilla).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92078860.
RA   Lawlor D.A., Warren E., Taylor P., Parham P.;
RT   "Gorilla class I major histocompatibility complex alleles: comparison
RT   to human and chimpanzee class I.";
RL   J. Exp. Med. 174:1491-1509(1991).
CC   -!- FUNCTION: INVOLVED IN THE PRESENTATION OF F
CC       THE IMMUNE SYSTEM.
CC   -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN
CC       MICROGLOBULIN).
CC   ---------- ---------- -- - ------ ------ -
CC   This SWISS-PROT entry is copyright. It is produced
CC   between the Swiss Institute of Bioinformatics an
CC   the European Bioinformatics Institute. There are
CC   use by non-profit institutions as long as its
CC   modified and this statement is not removed. Usage
CC   entities requires a license agreement (See http://
CC   or send an email to license@isb-sib.ch).
CC   ---------- ---- -- - ---------
DR   EMBL; X60251; CAA42804.1; -.
DR   PIR; JH0545; JH0545.
DR   HSSP; P30460; 1AGB.
DR   PFAM; PF00129; MHC_I; 1.
DR   PFAM; PF00047; ig; 1.
DR   PROSITE; PS00290; IG_MHC; FALSE_NEG.
KW   MHC I; Transmembrane; Glycoprotein; Signal.
FT   SIGNAL        1     24        BY SIMILARITY.
FT   CHAIN        25    366        CLASS I HISTOCOMPATIB
FT                                 GOGO-C0201 ALPHA CHAI
FT   DOMAIN       25    114        EXTRACELLULAR ALPHA 1
FT   DOMAIN      115    206        EXTRACELLULAR ALPHA 2
FT   DOMAIN      207    298        EXTRACELLULAR ALPHA 3
FT   DOMAIN      299    308        CONNECTING PEPTIDE.
FT   TRANSMEM    309    333
FT   DOMAIN      334    366        CYTOPLASMIC TAIL.
FT   DISULFID    125    188        BY SIMILARITY.
FT   DISULFID    227    283        BY SIMILARITY.
FT   CARBOHYD    110    110        BY SIMILARITY.
SQ   SEQUENCE   366 AA;  40954 MW;  05E159364C769E05 CRC64;

  Query Match             17.2%;  Score 89;  DB 1;  Leng
  Best Local Similarity   33.3%;  Pred. No. 6.75e-01;
  Matches    17;  Conservative   10;  Mismatches   12;

Db      30 LIVVLFVVGIAVFTILKSLIDKKRKRSAKRGESPY 67
           :|| :| || :||: :      :  :: |:: |  | || ::
Qy      18 VPIGSIYVVDIAVFITLD-DATDPSIQWEVSRKGHPGGFCS 57


RESULT   11
ID   1C03_GORGO     STANDARD;      PRT:   366 AA.
AC   P30386;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   01-APR-1993 (Rel. 25, Last annotation update)
DE   CLASS I HISTOCOMPATIBILITY ANTIGEN, GOGO-C0202 A
OS   Gorilla gorilla gorilla (Lowland gorilla).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominid
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92078860.
RA   Lawlor D.A., Warren E., Taylor P., Parham P.;
RT   "Gorilla class I major histocompatibility complex
RT   to human and chimpanzee class I.";
RL   J. Exp. Med. 174:1491-1509(1991).
CC   -!- FUNCTION: INVOLVED IN THE PRESENTATION OF F
CC       THE IMMUNE SYSTEM.
CC   -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN
CC       MICROGLOBULIN).
CC   ---------- -- -- ------ -- -
CC   This SWISS-PROT entry is copyright. It is produced
CC   between the Swiss Institute of Bioinformatics an
CC   the European Bioinformatics Institute. There are
CC   use by non-profit institutions as long as its
CC   modified and this statement is not removed. Usage
CC   entities requires a license agreement (See http:/
```

```
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X60249; CAA42801.1; -.
DR   PIR; JH0546; JH0546.
DR   HSSP; P03989; 1BSA.
DR   PFAM; PF00129; MHC_1; 1.
DR   PFAM; PF00047; ig; 1.
DR   PROSITE; PS00290; IG_MHC; FALSE_NEG.
KW   MHC I; Transmembrane; Glycoprotein; Signal.
FT   SIGNAL       1     24       BY SIMILARITY.
FT   CHAIN       25    366       CLASS I HISTOCOMPATIBILITY ANTIGEN,
FT                               GOGO-C0202 ALPHA CHAIN.
FT   DOMAIN      25    114       EXTRACELLULAR ALPHA-1.
FT   DOMAIN     115    206       EXTRACELLULAR ALPHA-2.
FT   DOMAIN     207    298       EXTRACELLULAR ALPHA-3.
FT   DOMAIN     299    308       CONNECTING PEPTIDE.
FT   TRANSMEM   309    332
FT   DOMAIN     333    366       CYTOPLASMIC TAIL.
FT   DISULFID   125    188       BY SIMILARITY.
FT   DISULFID   227    283       BY SIMILARITY.
FT   CARBOHYD   110    110       BY SIMILARITY.
SQ   SEQUENCE   366 AA;  40782 MW;  11CCAD1F6091831B CRC64;

  Query Match           17.0%;  Score 89;  DB 1;  Length 366;
  Best Local Similarity  39.5%;  Pred. No. 6.75e-01;
  Matches   17;  Conservative   10;  Mismatches  12;  Indels   4;  Gaps   3;

Db     306 IPIVGI-VVGLAVLVVLAVLGAVVIAMMCRRKSS-GGKGGSCS 346
           :|| :| || :||:: |    |:  :: |:: |  | ||| ||
Qy      18 VPIGSIYVVDIAVFITLD--DAIFPSIGCWKVSRKGDKGGECS 58


RESULT   12
ID   1C04_GORGO     STANDARD;      PRT;   366 AA.
AC   P30387;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   01-APR-1993 (Rel. 25, Last annotation update)
DE   CLASS I HISTOCOMPATIBILITY ANTIGEN, GOGO-C0203 ALPHA CHAIN PRECURSOR.
OS   Gorilla gorilla gorilla (Lowland gorilla).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92078860.
RA   Lawlor D.A., Warren E., Taylor P., Parham P.;
RT   "Gorilla class I major histocompatibility complex alleles: comparison
RT   to human and chimpanzee class I.";
RL   J. Exp. Med. 174:1491-1509(1991).
CC   -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC       THE IMMUNE SYSTEM.
CC   -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2
CC       MICROGLOBULIN).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X60248; CAA42800.1; -.
DR   PIR; JH0547; JH0547.
DR   HSSP; P30460; 1AGB.
DR   PFAM; PF00129; MHC_1; 1.
DR   PFAM; PF00047; ig; 1.
DR   PROSITE; PS00290; IG_MHC; FALSE_NEG.
KW   MHC I; Transmembrane; Glycoprotein; Signal.
FT   SIGNAL       1     24       BY SIMILARITY.
FT   CHAIN       25    366       CLASS I HISTOCOMPATIBILITY ANTIGEN,
FT                               GOGO-C0203 ALPHA CHAIN.
FT   DOMAIN      25    114       EXTRACELLULAR ALPHA-1.
FT   DOMAIN     115    206       EXTRACELLULAR ALPHA-2.
FT   DOMAIN     207    298       EXTRACELLULAR ALPHA-3.
FT   DOMAIN     299    308       CONNECTING PEPTIDE.
FT   TRANSMEM   309    332
FT   DOMAIN     333    366       CYTOPLASMIC TAIL.
FT   DISULFID   125    188       BY SIMILARITY.
FT   DISULFID   227    283       BY SIMILARITY.
FT   CARBOHYD   110    110       BY SIMILARITY.
SQ   SEQUENCE   366 AA;  40970 MW;  EE962C8189CAC001 CRC64;

  Query Match           17.0%;  Score 89;  DB 1;  Length 366;
  Best Local Similarity  39.5%;  Pred. No. 6.75e-01;
  Matches   17;  Conservative   10;  Mismatches  12;  Indels   4;  Gaps   3;

Db     306 IPIVGI-VVGLAVLVVLAVLGAVVTAMMCRRKSS-GGKGGSCS 346
           :|| :| || :||:: |    |:  :: |:: |  | ||| ||
Qy      18 VPIGSIYVVDIAVFITLD--DAIFPSIGCWKVSRKGDKGGECS 58


RESULT   13
ID   NAH2_RABIT     STANDARD;      PRT;   809 AA.
AC   P50482;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   SODIUM/HYDROGEN EXCHANGER 2 (NA(+)/H(+) EXCHANGER 2) (NHE 2).
GN   SLC9A2 OR NHE2.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN NEW ZEALAND WHITE;
RX   MEDLINE; 93280159.
RA   Tse C.-M., Levine S.A., Yun C.H., Montrose M.H., Little P.J.,
RA   Pouyssegur J., Donowitz M.;
RT   "Cloning and expression of a rabbit cDNA encoding a serum-activated
RT   ethylisopropylamiloride-resistant epithelial Na+/H+ exchanger isoform
RT   (NHE 2).";
RL   J. Biol. Chem. 268:11917-11924(1993).
CC   -!- FUNCTION: INVOLVED IN PH REGULATION TO ELIMINATE ACIDS GENERATED
CC       BY ACTIVE METABOLISM OR TO COUNTER ADVERSE ENVIRONMENTAL
CC       CONDITIONS. MAJOR PROTON EXTRUDING SYSTEM DRIVEN BY THE INWARD
CC       SODIUM ION CHEMICAL GRADIENT. PLAYS AN IMPORTANT ROLE IN SIGNAL
CC       TRANSDUCTION.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: HIGH LEVELS IN INTESTINE AND KIDNEY.
CC   -!- PTM: PHOSPHORYLATED (POSSIBLE).
CC   -!- SIMILARITY: BELONGS TO THE NA(+)/H(+) EXCHANGER FAMILY.
CC   -!- CAUTION: THE NUMBER, LOCALIZATION AND DENOMINATION OF HYDROPHOBIC
CC       DOMAINS IN THE NA(+)/H(+) EXCHANGERS VARY AMONG AUTHORS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L13733; -; NOT_ANNOTATED_CDS.
DR   PFAM; PF00999; Na_H_Exchanger; 1.
DR   PRINTS; PR01084; NAHEXCHNGR.
DR   PRINTS; PR01086; NAHEXCHNGR2.
KW   Transmembrane; Glycoprotein; Sodium transport; Transport; Symport;
KW   Multigene family; Phosphorylation.
FT   DOMAIN       1     16       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    17     37       A (M1) (POTENTIAL).
FT   DOMAIN      38     79       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    80    100       B (M2) (POTENTIAL).
FT   DOMAIN     101    106       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   107    127       C (M3) (POTENTIAL).
FT   DOMAIN     128    138       EXTRACELLULAR (POTENTIAL).
```

```
FT   TRANSMEM     139     159       D (M4) (POTENTIAL).
FT   DOMAIN       160     168       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     169     189       E (M5) (POTENTIAL).
FT   DOMAIN       190     208       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     209     229       F (M5A) (POTENTIAL).
FT   DOMAIN       230     236       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     237     257       G (M5B) (POTENTIAL).
FT   DOMAIN       258     277       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     278     298       H (M6) (POTENTIAL).
FT   DOMAIN       299     307       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     308     328       I (M7) (POTENTIAL).
FT   DOMAIN       329     360       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     361     381       J (M8) (POTENTIAL).
FT   DOMAIN       382     391       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     392     412       K (M9) (POTENTIAL).
FT   DOMAIN       413     429       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     430     450       L (M10) (POTENTIAL).
FT   DOMAIN       451     458       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     459     479       M (M11) (POTENTIAL).
FT   DOMAIN       480     809       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD     350     350       POTENTIAL.
SQ   SEQUENCE   809 AA;  90744 MW;  DBD00B4544D87A6 CRC64;

  Query Match           17.0%;  Score 89;  DB 1;  Length 809;
  Best Local Similarity  31.0%;  Pred. No. 6.75e-01;
  Matches   18;  Conservative   13;  Mismatches  25;  Indels   2;  Gaps   2;

Db    453 VFPRKKLFIT-AAIVVIFFIVFI-LGIIIRPLVEFLDVKRSNKKQQAVSFFIHCREFD 508
          :|    ||:  ::| |: ::||| |  :| | :    | |   |    | ::|| :|:
Qy     10 IFLNSYLFVPIGSIYVVDIAVFITLDDAIFPSIGCWKVSRKGDKGGECSLDHCGLFQ 67


RESULT   14
ID   NAH2_RAT       STANDARD;      PRT;   813 AA.
AC   P48763;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   SODIUM/HYDROGEN EXCHANGER 2 (NA(+)/H(+) EXCHANGER 2) (NHE-2) (H7).
GN   SLC9A2 OR NHE2.
OS   Rattus norvegicus (Rat)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=STOMACH;
RX   MEDLINE; 93280160.
RA   Wang Z., Orlowski J., Shull G.E.;
RT   "Primary structure and functional expression of a novel
RT   gastrointestinal isoform of the rat Na/H exchanger.";
RL   J. Biol. Chem. 268:11925-11928(1993).
RN   [2]
RP   SEQUENCE OF 117-813 FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=SMALL INTESTINE;
RX   MEDLINE; 93248205.
RA   Collins J.F., Honda T., Knobel S., Bulus N.M., Conary J.,
RA   Dubois R., Chisham F.K.;
RT   "Molecular cloning, sequencing, tissue distribution, and functional
RT   expression of a Na+/H+ exchanger (NHE-2).";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:3938-3942(1993).
CC   -!- FUNCTION: INVOLVED IN PH REGULATION TO ELIMINATE ACIDS GENERATED
CC       BY ACTIVE METABOLISM OR TO COUNTER ADVERSE ENVIRONMENTAL
CC       CONDITIONS. MAJOR PROTON EXTRUDING SYSTEM DRIVEN BY THE INWARD
CC       SODIUM ION CHEMICAL GRADIENT. PLAYS AN IMPORTANT ROLE IN SIGNAL
CC       TRANSDUCTION.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- ALTERNATIVE PRODUCTS: TWO FORMS MAY BE PRODUCED BY ALTERNATIVE
CC       SPLICING.
CC   -!- TISSUE SPECIFICITY: PREDOMINANTLY IN SMALL INTESTINE, COLON, AND
CC       STOMACH, WITH MUCH LOWER LEVELS IN SKELETAL MUSCLE, KIDNEY, BRAIN,
CC       TESTIS, UTERUS, HEART, AND LUNG.
CC   -!- PTM: PHOSPHORYLATED (POSSIBLE).
CC   -!- SIMILARITY: BELONGS TO THE NA(+)/H(+) EXCHANGER FAMILY.
CC   -!- CAUTION: THE NUMBER, LOCALIZATION AND DENOMINA...
CC       DOMAINS IN THE NA(+)/H(+) EXCHANGERS VARY AMON...
CC   ----------...
CC   This SWISS-PROT entry is copyright. It is produced...
CC   between the Swiss Institute of Bioinformatics an...
CC   the European Bioinformatics Institute. There are ...
CC   use by non-profit institutions as long as it's ...
CC   modified and this statement is not removed. Usage...
CC   entities requires a license agreement (See http://...
CC   or send an email to license@isb-sib.ch).
CC   ----------...
DR   EMBL; L11236; AAA72...
DR   EMBL; L11004; AAA754...
DR   PFAM; PF00999; Na_H_Exchanger; 1.
DR   PRINTS; PR01084; NAHEXCHNGR.
DR   PRINTS; PR01086; NAHEXCHNGR2.
KW   Transmembrane; Glycoprotein; Sodium transport; Ion...
KW   Multigene family; Phosphorylation; Alternative spl...
FT   DOMAIN         1      13       CYTOPLASMIC (POTENTIA...
FT   DOMAIN        14      34       A (M1) HYDROPHOBIC.
FT   DOMAIN        35      80       EXTRACELLULAR (POTENTIA...
FT   DOMAIN        81     101       B (M2) HYDROPHOBIC.
FT   DOMAIN       102     107       CYTOPLASMIC (POTENTIA...
FT   TRANSMEM     108     128       C (M3) (POTENTIAL).
FT   DOMAIN       129     139       EXTRACELLULAR (POTENTI...
FT   TRANSMEM     140     160       D (M4) (POTENTIAL).
FT   DOMAIN       161     169       CYTOPLASMIC (POTENTIA...
FT   TRANSMEM     170     190       E (M5) (POTENTIAL).
FT   DOMAIN       191     209       EXTRACELLULAR (POTENTI...
FT   TRANSMEM     210     230       F (M5A) (POTENTIAL).
FT   DOMAIN       231     237       CYTOPLASMIC (POTENTIA...
FT   TRANSMEM     238     258       G (M5B) (POTENTIAL).
FT   DOMAIN       259     278       EXTRACELLULAR (POTENTI...
FT   TRANSMEM     279     299       H (M6) (POTENTIAL).
FT   DOMAIN       300     308       CYTOPLASMIC (POTENTIA...
FT   TRANSMEM     309     329       I (M7) (POTENTIAL).
FT   DOMAIN       330     361       EXTRACELLULAR (POTENTI...
FT   TRANSMEM     362     382       J (M8) (POTENTIAL).
FT   DOMAIN       383     392       CYTOPLASMIC (POTENTIA...
FT   TRANSMEM     393     413       K (M9) (POTENTIAL).
FT   DOMAIN       414     430       EXTRACELLULAR (POTENTI...
FT   DOMAIN       431     451       L. HYDROPHOBIC.
FT   DOMAIN       452     479       EXTRACELLULAR (POTENTI...
FT   TRANSMEM     460     480       M11 (POTENTIAL).
FT   DOMAIN       481     813       CYTOPLASMIC (POTENTIA...
FT   CARBOHYD     351     351       POTENTIAL.
FT   VARSPLIC       1     119       MISSING (IN SHORT IS...
FT   CONFLICT     786     786       V -> G (IN REF. 2).
SQ   SEQUENCE   813 AA;  914...2 MW;  2...27267D7085...CRC...

  Query Match           17.0%;  Score 89;  DB 1;  Length...
  Best Local Similarity  31.0%;  Pred. No. 6.75e-01;
  Matches   18;  Conservative   13;  Mismatches  25;  ...

Db    454 VFPRKKLFIT-AAIVVIFFIVFI-LGIIIRPLVEFLDVKRSNKKQQ...
          :|    ||:  ::| |: ::||| |  :| | :    | |   |...
Qy     10 IFLNSYLFVPIGSIYVVDIAVFITLDDAIFPSIGCWKVSRKGDKGG...


RESULT   15
ID   SYS_BORBU       STANDARD;      PRT:   425 AA.
AC   O51244;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   SERYL-TRNA SYNTHETASE (EC 6.1.1.11) (SERINE--TRNA L...
GN   SERS OR BB0226.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 35210 / B31;
RX   MEDLINE; 98065943.
```

```
RA   Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA   Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA   Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA   Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA   van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA   Utterback T., Watthey L., McDonald L., Artiach P., Bowman C.,
RA   Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA   Smith H.O., Venter J.C.;
RT   "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT   burgdorferi.";
RL   Nature 390:580-586(1997).
CC   -!- CATALYTIC ACTIVITY: ATP + L-SERINE + TRNA(SER) = AMP +
CC       PYROPHOSPHATE + L-SERYL-TRNA(SER).
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE001133; AAC66622.1; -.
DR   TIGR; BB0226; -.
DR   PFAM; PF00587; tRNA-synt_2b; 1.
DR   PRINTS; PR00981; TRNASYNTHSER.
DR   PROSITE; PS00179; AA_TRNA_LIGASE_II_1; 1.
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; FALSE_NEG.
KW   Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.
SQ   SEQUENCE   425 AA;  48707 MW;  9AE5CEBB429B5A3B CRC64;

  Query Match           16.9%;  Score 88;  DB 1;  Length 425;
  Best Local Similarity  25.7%;  Pred. No. 9.17e-01;
  Matches   18;  Conservative   18;  Mismatches  29;  Indels   5;  Gaps   4;

Db     292 SGVIHEEPLSIQP-IFTPTFIPYEVLNGSF FLGSPAYEEKYDLEAWMPCKMRGGYGHV 350
           :|:|:::|: :  :|  :  | |::|  |  |  : :    |  |: :| |  | ||
Qy       2 AGLINEDFTFLNSYLFVPIGSIY-VVDIAVFTTLDDAIFP--SIGCWKVSRKGDKG-GEC 57

Db     351 TSTSNCTDYQ 360
           :   :|  :|
Qy      58 SLDFHCGLFQ 67


Search completed: Fri Aug 25 17:57:20 2000
Job time : 18 secs.
```

MPsrch (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 25 17:56:25 2000; MasPar time 6.35 Seconds
505.405 Million cell updates/sec

Tabular output not generated.

Title: >US-09-257-585-5
Description: (1-68) from US09257585.pep
Perfect Score: 522
Sequence: 1 MACLINEDFIFLNSYILVPI.....RKGDKGGECSLDFHCGLFQL 68

Scoring table: PAM 150
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir64
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 48.943; Variance 87.059; scale 0.447

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	99	19.0	873	2	S26690	glycoprotein B - infe	2.04e-01
2	99	19.0	883	1	VGBEIL	glycoprotein B precur	2.04e-01
3	94	18.0	864	2	I46604	MHC PD14 transplantat	8.40e-01
4	94	18.0	866	2	I46603	MHC PD14a transplanta	8.40e-01
5	94	18.0	883	1	VGBEIS	glycoprotein B precur	8.40e-01
6	93	17.8	138	2	S30708	hypothetical 15.4K pr	1.11e+00
7	93	17.8	344	2	S14409	membrane protein, 37K	1.11e+00
8	93	17.8	366	2	A60369	MHC class I histocomp	1.11e+00
9	92	17.6	437	2	S34959	NADH dehydrogenase (u	1.46e+00
10	92	17.6	1461	2	C71403	hypothetical protein	1.46e+00
11	91	17.4	167	2	T16454	hypothetical protein	1.92e+00
12	91	17.4	411	2	S73218	preprotein translocas	1.92e+00
13	91	17.4	527	2	C64626	alginate O acetylatio	1.92e+00
14	91	17.4	537	2	G72576	hypothetical protein	1.92e+00
15	90	17.2	283	2	B70171	conserved hypothetica	2.53e+00
16	89	17.0	366	2	JH0545	class I histocompatib	3.31e+00
17	89	17.0	366	2	JH0547	class I histocompatib	3.31e+00
18	89	17.0	366	2	JH0546	class I histocompatib	3.31e+00
19	89	17.0	809	2	A45747	Na+/H+-exchanging pro	3.31e+00
20	89	17.0	813	2	A46748	Na+/H+-exchanging pro	3.31e+00
21	88	16.9	425	2	B70129	serine tRNA ligase (	4.33e+00
22	88	16.9	428	2	H71044	hypothetical protein	4.33e+00
23	88	16.9	877	1	A25962	mitosis inhibitor wee	4.33e+00
24	87	16.7	[illegible]	2	A43409	T-cell sur	
25	87	16.7	[illegible]	2	JH0544	class I hi	
26	87	16.7	[illegible]	2	C29189	nitedipine	
27	87	16.7	[illegible]	2	[illegible]	probable ar	
28	87	16.7	[illegible]	2	[illegible]	S5kl prote	
29	86	16.5	[illegible]	2	[illegible]	Mo147 homo	
30	85	16.3	101	2	A26606	NADH dehyd	
31	85	16.3	174	1	[illegible]	NADH dehyd	
32	85	16.3	[illegible]	2	[illegible]	hypothetic	
33	85	16.3	412	2	[illegible]	hypothetic	
34	85	16.3	[illegible]	2	[illegible]	probable ne	
35	84	16.1	471	2	A23451	NADH dehyd	
36	84	16.1	480	2	S15808	hypothetic	
37	84	16.1	489	2	C22845	NADH dehyd	
38	84	16.1	973	2	[illegible]	hypothetic	
39	84	16.1	1076	2	S[illegible]	1,3-beta-g	
40	83	15.9	[illegible]	2	[illegible]	hypothetic	
41	83	15.9	[illegible]	2	[illegible]	sugar phos	
42	83	15.9	1084	2	[illegible]	cellulose	
43	83	15.9	1084	2	[illegible]	cellulose	
44	83	15.9	1102	2	S34772	C29B4.4 pr	
45	83	15.9	1805	2	S[illegible]	1,3-beta-g	

ALIGNMENTS

RESULT 1
ENTRY S26690 #type complete
TITLE glycoprotein B - infectious laryngotra
ORGANISM #formal_name Infectious laryngotracheit
DATE 13-Jun-1995 #sequence_revision 13-Jun
07-May-1999
ACCESSIONS S26690
REFERENCE S26690
#authors Poulsen, D.J.; Adams Burton, C.K.; [illegible]
S.J.; Keeler Jr., C.L.
#journal Virus Genes (1991) 5:335-347
#title Identification of the infectious laryn
glycoprotein gB gene by the polymera
#cross-references MUID:92186
#accession S26690
##status preliminary
##molecule_type DNA
##residues 1-873 ##label pol
##cross-references EMBL:X56984
CLASSIFICATION #superfamily herpesvirus glycoprotein
KEYWORDS glycoprotein; transmembrane protein;
SUMMARY #length 873 #molecular-weight 99124

Query Match 19.0%; Score 99; DB 2; Lengt
Best Local Similarity 24.3%; Pred. No. 2.04e-01;
Matches 9; Conservative 10; Mismatches 11;

Db 607 LLGENYLLYEDYTFVKQVNAC[illegible]PVS[illegible]PINLNATIL 664
[illegible]
Qy 4 LINEDFIFLNSYILV rlaSIYVVCIAVFITLDDAIF 39

RESULT 2
ENTRY VGBEIL #type complete
TITLE glycoprotein B precursor - Infectious
(strain Thorne V882)
ORGANISM #formal_name Infectious laryngotracheit
#note host Gallus gallus (chicken)
DATE 31-Dec-1991 #sequence_revision 31-Dec
16-Jul-1999
ACCESSIONS A38478; JQ0951
REFERENCE A38478
#authors Griffin, A.M.
#journal J. Gen. Virol. (1991) 72:393-398
#title The nucleotide sequence of the glycopr
infectious laryngotracheitis virus: a
evolutionary relationship to the hom

```
                herpesviruses.
   #cross-references MUID:91132136
   #accession     A38478
      ##molecule_type DNA
      ##residues       1-883 ##label GRI
      ##cross-references GB:M30818, NID:g221905, PIDN:BAA00693.1,
                        PID:d1001157; PID:g221907
COMMENT     This transmembrane protein is essential for viral infectivity.
GENETICS
   #gene           gB
CLASSIFICATION    #superfamily herpesvirus gI glycoprotein B
KEYWORDS          glycoprotein; transmembrane protein
FEATURE
   1-26                #domain signal sequence #status predicted #label SIG\
   27-883              #product glycoprotein #status predicted #label GCP\
   725-744             #domain transmembrane #status predicted #label TM1\
   751-771             #domain transmembrane #status predicted #label TM2\
   102,121,211,262,
   360,579,635,649     #binding_site carbohydrate (Asn) (covalent) #status
                         predicted
SUMMARY           #length 883  #molecular-weight 100160  #checksum 4968

  Query Match             19.0%;  Score 99;  DB 1;  Length 883;
  Best Local Similarity  24.3%;  Pred. No. 2.04e-01;
  Matches     9;  Conservative    16;  Mismatches  11;  Indels    1;  Gaps    1;

Db     617 LLGENYLLYEDYTFVRQVNASEIEEVSIFINLNATIL 653
           |: |:::: : | ||   : :   : :::::|| |: :|:
Qy       4 LINEDFIFLNSYLFV-PIGSIYVVDIAVFITLDDAIF 39


RESULT    3
ENTRY             I46604      #type complete
TITLE             MHC PD14 transplantation antigen - pig
ORGANISM          #formal_name Sus scrofa domestica #common_name domestic pig
DATE              21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change
                    23-Jul-1999
ACCESSIONS        I46604
REFERENCE         I46603
   #authors       Satz, L.L.; Wang, L.
   #journal       J. Immunol. (1985) 135:2167-2175
   #title         Structure and expression of two porcine genomic clones
                    encoding class I MHC antigens.
   #cross-references MUID:85262925
   #accession     I46604
      ##status         preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues       1-364 ##label SAT
      ##cross-references GB:M21058; NID:g164556; PIDN:AAA31079.1; PID:g164558
GENETICS
   #introns        25:1, 115:1, 197:1, 259:1, 306:1, 347:1, 364:1
CLASSIFICATION    #superfamily class I histocompatibility antigen;
                    immunoglobulin homology
SUMMARY           #length 364  #molecular-weight 40605  #checksum 4699

  Query Match             18.0%;  Score 94;  DB 2;  Length 364;
  Best Local Similarity  34.2%;  Pred. No. 8.40e-01;
  Matches    13;  Conservative    12;  Mismatches  11;  Indels    2;  Gaps    2;

Db     305 VPIVGI-IVGL-VLVLVAGAVVAGVVIWRKKRSGEKGG 340
           ||| :| :| : |:: :  |: ::  |: | |:|||
Qy      18 VPIGSIYVVDIAVFITLDDAIFPSIGCWKVSRKGDKGG 55


RESULT    4
ENTRY             I46603      #type complete
TITLE             MHC PD14a transplantation antigen - pig
ORGANISM          #formal_name Sus scrofa domestica #common_name domestic pig
DATE              21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change
                    23-Jul-1999
ACCESSIONS        I46603
REFERENCE         I46603
   #authors       Satz, L.L.; Wang, L.
   #journal       J. Immunol. (1985) 135:2167-2175
   #title         Structure and expression of two porcine genomic clones
                    encoding class I MHC antigens.
   #cross-references MUID:85262925
   #accession     I46603
      ##status         preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues       1-366 ##label SAT
      ##cross-references GB:M21058; NID:g164556; PIDN:AAA31079.1; PID:g164558
GENETICS
   #introns        25:1, 115:1, 197:1, 259:1, 306:1, 347:1, 364:1
CLASSIFICATION    #superfamily class I histocompatibility antigen;
                    immunoglobulin homology
SUMMARY           #length 366  #molecular-weight 40796  #checksum 7412

  Query Match             18.0%;  Score 94;  DB 2;  Length 366;
  Best Local Similarity  34.2%;  Pred. No. 8.40e-01;
  Matches    13;  Conservative    12;  Mismatches  11;  Indels    2;  Gaps    2;

Db     305 VPIVGI-IVGL-VLVLVAGAVVAGVVIWRKKRSGEKGG 340
           ||| :| :| : |:: :  |: ::  |: | |:|||
Qy      18 VPIGSIYVVDIAVFITLDDAIFPSIGCWKVSPKGDKGG 55


RESULT    5
ENTRY             VGBEIS      #type complete
TITLE             glycoprotein B precursor - infectious laryngotracheitis virus
                    (strain SA-2)
ORGANISM          #formal_name infectious laryngotracheitis virus
DATE              30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
                    09-Sep-1994
ACCESSIONS        A40567
REFERENCE         A40567
   #authors       Kongsuwan, K.; Prideaux, C.T.; Johnson, M.A.; Sheppard, M.;
                    Fahey, K.J.
   #journal       Virology (1991) 184:404-410
   #title         Nucleotide sequence of the gene encoding infectious
                    laryngotracheitis virus glycoprotein B.
   #cross-references MUID:91335774
   #accession     A40567
      ##molecule_type DNA
      ##residues       1-883 ##label KON
      ##cross-references GB:M64927
CLASSIFICATION    #superfamily herpesvirus glycoprotein B
KEYWORDS          glycoprotein; transmembrane protein
FEATURE
   1-26                #domain signal sequence #status predicted #label SIG\
   27-883              #product glycoprotein B #status predicted #label GPB\
   725-744             #domain transmembrane #status predicted #label TM1\
   751-771             #domain transmembrane #status predicted #label TM2\
   102,121,211,262,
   360,579,635,649     #binding_site carbohydrate (Asn) (covalent) #status
                         predicted
SUMMARY           #length 883  #molecular-weight 100112  #checksum 6911

  Query Match             18.0%;  Score 94;  DB 1;  Length 883;
  Best Local Similarity  24.3%;  Pred. No. 8.40e-01;
  Matches     9;  Conservative    15;  Mismatches  12;  Indels    1;  Gaps    1;

Db     617 LLGENYLLYEDYTFVRQVNASEIEEVSTFINLNATIL 653
           |: |:::: : | ||   : :   : ::: || |: :|:
Qy       4 LINEDFIFLNSYLFV-PIGSIYVVDIAVFITLDDAIF 39


RESULT    6
ENTRY             S30708      #type complete
TITLE             hypothetical 15.8K protein (corA-rarD intergenic region) -
                    Escherichia coli
ORGANISM          #formal_name Escherichia coli
DATE              02-Dec-1993 #sequence_revision 01-Sep-1995 #text_change
                    14-Nov-1997
ACCESSIONS        S30708; S30745; C65186
REFERENCE         S30660
```

```
   #authors       Daniels, D.L.; Plunkett III, G.; Burland, V.; Blattner, F.R.
   #journal       Science (1992) 257:771-778
   #title         Analysis of the Escherichia coli genome: DNA sequence of the
                    region from 84.5 to 86.5 minutes.
   #cross-references MUID:92358234
   #accession     S30708
      ##status          preliminary; nucleic acid sequence not shown;
                         translation not shown
      ##molecule_type DNA
      ##residues       1-138 ##label DAN
      ##cross-references EMBL:M87049; NID:g836656; PID:g148217
      ##note           the nucleotide sequence was submitted to the EMBL Data
                         Library, November 1992
REFERENCE         S30742
   #authors       Ohmori, H.
   #submission    submitted to the EMBL Data Library, January 1993
   #description   Physical map of the corA region of the E.coli chromosome.
   #accession     S30745
      ##status          preliminary
      ##molecule_type DNA
      ##residues       1-138 ##label OHM
      ##cross-references EMBL:L02122; NID:g145575; PID:g145579
REFERENCE         A64720
   #authors       Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
                    Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
                    Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
                    Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
                    Y.
   #journal       Science (1997) 277:1453-1462
   #title         The complete genome sequence of Escherichia coli K-12.
   #cross-references MUID:97426617
   #accession     C65186
      ##status          preliminary; nucleic acid sequence not shown;
                         translation not shown
      ##molecule_type DNA
      ##residues       1-138 ##label BLAT
      ##cross-references GB:AE000457; GB:U00096; NID:g2367294; PID:g1790251;
                         UWGP:b3818
      ##experimental_source strain K-12, substrain MG1655
GENETICS
   #gene          yiqG
   #map_position 86 min
SUMMARY           #length 138  #molecular-weight 15794  #checksum 1118

  Query Match             17.8%;  Score 93;  DB 2;  Length 138;
  Best Local Similarity   28.1%;  Pred. No. 1.1e+00;
  Matches    9;  Conservative    12;  Mismatches   10;  Indels    1;  Gaps    1;

Db      76 LRDSGFIYIKIYVFGIL-AVYIISIIIMIAED 106
           | :  |||:::|:|  : :::: |   |: |
Qy       4 LINEDFIFLNSYLFVPIGSIYVVDIAVFITLD 35


RESULT    7
ENTRY             S14409      #type complete
TITLE             membrane protein, 37K, precursor, chloroplast inner envelope
                    - spinach
ORGANISM          #formal_name Spinacia oleracea #common_name spinach
DATE              21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change
                    01-May-1998
ACCESSIONS        S14409; S16986
REFERENCE         S14409
   #authors       Dreses-Werringloer, U.; Fischer, K.; Wachter, E.; Link, T.A.;
                    Fluegge, U.I.
   #journal       Eur. J. Biochem. (1991) 195:361-368
   #title         cDNA sequence and deduced amino acid sequence of the
                    precursor of the 37-kDa inner envelope membrane polypeptide
                    from spinach chloroplasts. Its transit peptide contains an
                    amphiphilic alpha-helix as the only detectable structural
                    element.
   #cross-references MUID:91146572
   #accession     S14409
      ##molecule_type mRNA
      ##residues       1-344 ##label ESE
      ##cross-references EMBL:X[illegible]; NID:g21227; PID:g
REFERENCE         S16986
   #authors       Block, M.A.; [illegible]; Douce, R.
   #journal       FEBS Lett. (1991) 287:167-170
   #title         Purification and characterization of [illegible]
                    envelope protein.
   #cross-references MUID:[illegible]
   #accession     S16986
      ##molecule_type protein
      ##residues       1-[illegible]
GENETICS
   #genome        nuclear
CLASSIFICATION    #superfamily [illegible] homology
KEYWORDS          blocked amino end; chloroplast
FEATURE
   1-21                #domain: transit peptide (chloroplast)
                         #label: TNE\
   22-344              #product membrane protein, 37K #status
                         MA:\
   118-219             #domain: [illegible] homology #label: BI
SUMMARY           #length 344  #molecular-weight [illegible]

  Query Match             [illegible]%;  Score 93;  DB 2;  Length
  Best Local Similarity   [illegible]%;  Pred. No. 1.1[illegible]
  Matches    10;  Conservative    13;  Mismatches    9;  [illegible]

Db     308 LVFLYRFILGALASIYYVLVEIYMWIKDKIFP 339
           :|||  ::: :::| | | ::::: :  .|||
Qy       9 FIFLNSYLFVPIGSIYVVDIAVFITLDAIFP 40


RESULT    8
ENTRY             A60369      #type complete
TITLE             MHC class I histocompatibility antigen
                    precursor - pig
ORGANISM          #formal_name Sus scrofa domestica #[illegible]
DATE              03-Mar-1993 #sequence_revision 14-Mar-
                    17-Mar-1999
ACCESSIONS        A60369
REFERENCE         A60369
   #authors       Frels, W.I.; Bodmer, C.; Goldina, H
                    Rudikoff, S.; Singer, D.S.
   #journal       New Biol. (1990) 2:1024-1033
   #title         Expression of a class I MHC transgene
                    tissue-specific negative regulatory
   #cross-references MUID:91329292
   #accession     A60369
      ##molecule_type DNA
      ##residues       1-368 ##label FRE
      ##note           the authors failed to translate
                         at positions 22, 118, 301, 338
                         exons may not be fully translat
                         nucleotides, as shown, have corres
                         sequences
CLASSIFICATION    #superfamily class I histocompatibili
                    immunoglobulin homology
KEYWORDS          glycoprotein; heterodimer; transmembr
FEATURE
   222-287             #domain immunoglobulin homology #
SUMMARY           #length 368  #molecular-weight 41[illegible]

  Query Match             [illegible]%;  Score 93;  DB 2;  Leng
  Best Local Similarity   34.2%;  Pred. No. 1.1[illegible];
  Matches    13;  Conservative    11;  Mismatches   12;  [illegible]

Db     297 VFIVGI IVGL VLVIVAGAVVF YVV.WRKKRSGEKGG 42
           ||| :| :| |:: :  |:  ::  |:  | |:|||
Qy      16 VPIGSIYVVDIAVFIIDDA[illegible]WKVSPEGDKGG 57


RESULT    9
ENTRY             S34959      #type complete
```

```
TITLE             NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 4 -
                    Crithidia oncopelti mitochondrion (SGC6)
ORGANISM          #formal_name mitochondrion Crithidia oncopelti
DATE              30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
                    11-Jun-1999
ACCESSIONS        S34959
REFERENCE         S34958
   #authors       Muslow, D.A.; Horvath, A.; Gwang, H., K.; Kolesnikov, A.A.
   #submission    submitted to the EMBL Data Library, October 1990
   #accession     S34959
      ##molecule_type DNA
      ##residues       1-437 ##label MAS
      ##cross-references EMBL:X56015; NID:g12879; PIDN:CAA39491.1; PID:g12881
GENETICS
   #gene          ND4
   #genome        mitochondrion
   #genetic_code  SGC6
CLASSIFICATION    #superfamily NADH dehydrogenase (ubiquinone) chain 4
KEYWORDS          membrane-associated complex; mitochondrion; NAD; oxidative
                    phosphorylation; oxidoreductase; respiratory chain
SUMMARY           #length 437 #molecular-weight 52018 #checksum 2805

  Query Match           17.6%;  Score 92;  DB 2;  Length 437;
  Best Local Similarity  34.4%;  Pred. No. 1.46e+00;
  Matches   11;  Conservative   10;  Mismatches  10;  Indels   1;  Gaps   1;

Db     365 GLISLSFIYIICF-YVVIVLSVFVSSFIYTICL 395
           |||:  ||::  : :| : |::|  | ::| |
Qy       8 GLINEDFIFLNSYLFVPIGSIYVVDIAVFITL 44


RESULT   10
ENTRY             C71403      #type complete
TITLE             hypothetical protein  Arabidopsis thaliana
ORGANISM          #formal_name Arabidopsis thaliana #common_name mouse-ear
                    cress
   #variety       columbia
DATE              03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change
                    05-Dec-1998
ACCESSIONS        C71403
REFERENCE         A71400
   #authors       Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.;
                    Dean, C.; Bergkamp, R.; Dirkse, W.; Van Staveren, M.;
                    Stiekema, W.; Drost, L.; Ridley, P.; Hudson, S.A.; Patel,
                    K.; Murphy, G.; Piffanelli, P.; Wedler, H.; Wedler, E.;
                    Wambutt, R.; Weitzenegger, T.; Pohl, T.M.; Terryn, N.;
                    Gielen, J.; Villarroel, R.; De Clerck, R.; Van Montagu, M.;
                    Lecharny, A.; Aubora, S.; Gy, I.; Kreis, M.; Lao, N.;
                    Kavanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger,
                    M.; Schaefer, M.; Funk, B.; Mueller-Auer, S.; Silvey, M.;
                    James, R.; Montfort, A.; Pons, A.; Puigdomenech, P.; Douka,
                    A.; Voukelatou, E.; Milioni, D.; Hatzopoulos, P.;
                    Piravandi, E.; Obermaier, B.; Hilbert, H.; Duesterhoft, A.;
                    Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.;
                    Rechman, S.; Ansorge, W.; Cooke, R.; Berger, C.; Delseny,
                    M.; Voet, M.; Volckaert, G.; Mewes, H.W.; Klosterman, S.;
                    Schueller, C.; Chalwatzis, N.
   #journal       Nature (1998) 391:485-488
   #title         Analysis of 1.9 Mb of contiguous sequence from chromosome 4
                    of Arabidopsis thaliana.
   #cross-references MUID:98121113
   #accession     C71403
      ##status         preliminary; nucleic acid sequence not shown;
                         translation not shown
      ##molecule_type DNA
      ##residues       1-1361 ##label BEV
      ##cross-references GB:Z97335; NID:g2244747; PID:e327443; PID:g2244774
GENETICS
   #map_position 4COP9-4G3845
SUMMARY           #length 1361  #molecular-weight 151113  #checksum 9792

  Query Match           17.6%;  Score 92;  DB 2;  Length 1361;
  Best Local Similarity  42.4%;  Pred. No. 1.46e+00;
  Matches   11;  Conservative   13;  Mismatches  10;  Indels   0;  Gaps   0;

Db    1327 LPSSAFIPISSELILDISMFSALGNSLTAIKSVW 1360
           | |  |:||:|  ::||::| :| ::: :: : |
Qy      12 LNSYLFVPIGSIYVVDIAVFITLDDAIFPSIGCW 45


RESULT   11
ENTRY             T16454      #type complete
TITLE             hypothetical protein F53B3.5 - Caenorhabditis elegans
ORGANISM          #formal_name Caenorhabditis elegans
DATE              20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change
                    20-Sep-1999
ACCESSIONS        T16454
REFERENCE         Z18515
   #authors       Miller, N.
   #submission    submitted to the EMBL Data Library, November 1995
   #description   The sequence of C. elegans cosmid F53B3.
   #accession     T16454
      ##status         preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues       1-167 ##label MIL
      ##cross-references EMBL:U40414; NID:g1065474; PID:g1065480;
                         PIDN:AAA81409.1; CESP:F53B3.5
GENETICS
   #gene          CESP:F53B3.5
   #introns       42/1; 76/1; 108/3
SUMMARY           #length 167  #molecular-weight 18813  #checksum 6308

  Query Match           17.4%;  Score 91;  DB 2;  Length 167;
  Best Local Similarity  51.9%;  Pred. No. 1.92e+00;
  Matches   14;  Conservative    3;  Mismatches   8;  Indels   2;  Gaps   2;

Db      89 YVVGI-VFVVLS-FIVGLIGGWNKSKK 63
           ||| | ||: |   |   |||| |:|
Qy      24 YVVDIAVFITLDDAIFPSIGCWKVSRK 50


RESULT   12
ENTRY             S73218      #type complete
TITLE             preprotein translocase secY - red alga (Porphyra purpurea)
                    chloroplast
ORGANISM          #formal_name chloroplast Porphyra purpurea
DATE              19-Mar-1997 #sequence_revision 09-May-1997 #text_change
                    20-Sep-1999
ACCESSIONS        S73218
REFERENCE         S73108
   #authors       Reith, M.; Munholland, J.
   #journal       Plant Mol. Biol. Rep. (1995) 13:333-335
   #title         Complete nucleotide sequence of the Porphyra purpurea
                    chloroplast genome.
   #accession     S73218
      ##status         preliminary; nucleic acid sequence not shown;
                         translation not shown
      ##molecule_type DNA
      ##residues       1-411 ##label REI
      ##cross-references EMBL:U38804; NID:g1276652; PIDN:AAC08183.1;
                         PID:g1276763
      ##note           the nucleotide sequence was submitted to the EMBL Data
                         Library, October 1995
GENETICS
   #gene          secY
   #genome        chloroplast
CLASSIFICATION    #superfamily preprotein translocase secY
KEYWORDS          chloroplast; protein transport; transmembrane protein
SUMMARY           #length 411  #molecular-weight 45660  #checksum 38

  Query Match           17.4%;  Score 91;  DB 2;  Length 411;
  Best Local Similarity  33.3%;  Pred. No. 1.92e+00;
  Matches   14;  Conservative   12;  Mismatches  16;  Indels   0;  Gaps   0;

Db     269 LTQLTQNTFLLQVLYLFCPNGSLYLVLYSVLILFFSYFYTSI 310
           :: |  : |::   ||| | ||:|:|  :|:| :   :: ||
```

```
Qy       1 MAGLINEDFIFLNSYLFVPIGSIYVVDIAVFIILDDAIFPSI 42


RESULT   13
ENTRY            G64626      #type complete
TITLE            alginate O-acetylation protein - Helicobacter pylori (strain
                   26695)
ORGANISM         #formal_name Helicobacter pylori
DATE             09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change
                   10-Oct-1997
ACCESSIONS       G64626
REFERENCE        A64520
   #authors      Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.;
                   Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klenk,
                   H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush,
                   J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.;
                   Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.;
                   McKenney, K.; Fitzegerald, L.M.; Lee, N.; Adams, M.D.;
                   Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Utterback, T.R.;
                   Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.;
                   Fujii, C.; Bowman, C.; Watthey, L.; Wallin, E.; Hayes,
                   W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
                   C.M.; Venter, J.C.
   #journal      Nature (1997) 388:539-547
   #title        The complete genome sequence of the gastric pathogen
                   Helicobacter pylori.
   #cross-references MUID:97394467
   #accession    G64626
      ##status        preliminary; nucleic acid sequence not shown;
                       translation not shown
      ##molecule_type DNA
      ##residues      1-527 ##label HOM
      ##cross-references GB:AE000596; GB:AE000511; NID:g2313982; PID:g2313988;
                       TIGR:HP0855
SUMMARY          #length 527  #molecular-weight 61244  #checksum 6351

  Query Match           17.4%;  Score 91;  DB 2;  Length 527;
  Best Local Similarity  34.3%;  Pred. No. 1.92e+00;
  Matches   12;  Conservative    9;  Mismatches  13;  Indels   1;  Gaps   1;

Db     329 FLKEYLYIPLGGNRVKELIVYKNLI LVFLIGGFW 362
           ||: ||:::|:  | :: |:  |   :|   | |
Qy      11 FLNSYLFVPIGSIYVVDIAVFIILDDAIFPSIGGW 45


RESULT   14
ENTRY            G72576      #type complete
TITLE            hypothetical protein APE1895 - Aeropyrum pernix (strain K1)
ORGANISM         #formal_name Aeropyrum pernix
DATE             20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change
                   20-Aug-1999
ACCESSIONS       G72576
REFERENCE        A72450
   #authors      Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.;
                   Haikawa, Y.; Jin-no, K.; Takahashi, M.; Sekine, M.; Baba,
                   S.; Ankai, A.; Kosugi, H.; Hosoyama, A.; Fukui, S.; Nagai,
                   Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
                   Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.;
                   Kushida, N.; Oguchi, A.; Aoki, K.; Kubota, K.; Nakamura,
                   Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
   #journal      DNA Res. (1999) 6:83-101
   #title        Complete genome sequence of an aerobic hyper-thermophilic
                   Crenarchaeon, Aeropyrum pernix K1.
   #cross-references MUID:99310339
   #accession    G72576
      ##status        preliminary
      ##molecule_type DNA
      ##residues      1-537 ##label KAW
      ##cross-references DDBJ:AP000062; NID:g5105244; PIDN:BAA80900.1;
                       PID:d1044686; PID:g5105587
      ##experimental_source strain K1
GENETICS
   #gene         APE1895

SUMMARY          #length 537  #molecular-weight 59673

  Query Match           17.4%;  Score 91;  DB 2;  Length 537;
  Best Local Similarity  40.0%;  Pred. No. 1.92e+00;
  Matches   10;  Conservative    9;  Mismatches   5;  Indels

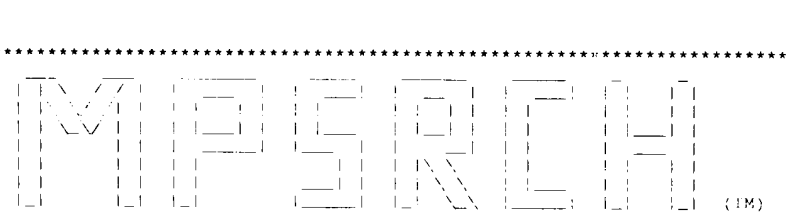
Db     499 YFYLALYLLLSAGSIFVAVLIIGVV 523
           :::|  |||:: |:|: :|:  :
Qy       9 FIFLNSYLFVPIGSIYV VLIAVFI 32


RESULT   15
ENTRY            E70171      #type complete
TITLE            conserved hypothetical integral membra
                   Lyme disease spirochete
ORGANISM         #formal name Borrelia burgdorferi #str
                   spirochete
DATE             13-Feb-1998 #sequence_revision 13-Feb
                   05-Jun-1998
ACCESSIONS       E70171
REFERENCE        A70100
   #authors      Fraser, C.M.; Casjens, S.; Huang, W.M
                   Clayton, R.; Lathigra, R.; White, O
                   Dodson, R.; Hickey, E.K.; Gwinn, M.
                   J.F.; Fleischmann, R.D.; Richardson
                   Kerlavage, A.R.; Quackenbush, J.; Sa
                   Vuqt, R.V.; Palmer, N.; Adams, M.D.;
                   J.; Utterback, T.; Watthey, L.; Mell
                   Bowman, C.; Garland, S.; Fujii, C.;
                   K.; Roberts, K.; Hatch, B.; Smith,
   #journal      Nature (1997)
   #title        Genomic sequence of a Lyme disease sp
                   burgdorferi.
   #cross-references MUID:98065943
   #accession    E70171
      ##status        preliminary; nucleic acid sequenc
                       translation not shown
      ##molecule_type DNA
      ##residues      1-283 ##label KLE
      ##cross-references GB:AE001158; GB:AE000785; NID;
                       TIGR:BB0574
      ##experimental_source strain B31
SUMMARY          #length 283  #molecular-weight 34177

  Query Match           17.2%;  Score 90;  DB 2;  Length
  Best Local Similarity  46.5%;  Pred. No. 2.5e+
  Matches   13;  Conservative    9;  Mismatches   4;

Db      96 EHYFFNLIKYLFALFLIIIIIYYLGVNIFLFLDNAFF 134
           | ::| |  ||| :  :  :  : : :: ||:||
Qy       7 EDFIF LNSYLF VI  IS IYVVDIAVFIILDAIF 39


Search completed: Fri Aug 25 17:59:34 2000
Job time : 20 secs.
```



Release 4.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 25 17:59:05 2000; MasPar time 3.02 Seconds
345.015 Million cell updates/sec

Tabular output not generated.

Title: >US-09-257-585-5
Description: (1-68) from US09257585.pep
Perfect Score: 522
Sequence: 1 MAGLINEDFTFLNGTLEVPI.....RKGDAGGPCSLDFHCGLFQL 68

Scoring table: PAM 150
Gap 11

Searched: 152433 seqs, 15829240 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:6_COMB 4:PCT_COMB 5:backfiles1

Statistics: Mean 25.744; Variance 107.081; scale 0.240

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description	Pred. No.
1	88	16.9	42	4	PCT-US93-0	Sequence 266, Applicat	5.82e+00
2	88	16.9	42	1	US-08-118-	Sequence 266, Applicat	5.82e+00
3	88	16.9	222	3	US-08-942-	Sequence 8, Applicatio	5.82e+00
4	87	16.7	182	2	US-08-284-	Sequence 25, Applicati	7.00e+00
5	87	16.7	182	4	PCT-US95-0	Sequence 25, Applicati	7.00e+00
6	87	16.7	182	2	US-08-417-	Sequence 25, Applicati	7.00e+00
7	87	16.7	182	4	PCT-US92-0	Sequence 25, Applicati	7.00e+00
8	84	16.1	46	1	US-08-118-	Sequence 268, Applicat	1.21e+01
9	84	16.1	46	4	PCT-US93-0	Sequence 268, Applicat	1.21e+01
10	84	16.1	471	2	US-08-928-	Sequence 20, Applicati	1.21e+01
11	84	16.1	1876	2	US-08-619-	Sequence 2, Applicatio	1.21e+01
12	83	15.9	1895	2	US-08-619-	Sequence 4, Applicatio	1.45e+01
13	81	15.5	269	2	US-07-857	Sequence 38, Applicati	2.85e+01
14	79	15.1	69	2	US-08-385-	Sequence 18, Applicati	2.99e+01
15	79	15.1	318	1	US-08-547-	Sequence 2, Applicatio	2.99e+01
16	77	14.8	419	2	US-08-270-	Sequence 2, Applicatio	4.27e+01
17	77	14.8	1294	2	US-08-819-	Sequence 3, Applicatio	4.27e+01
18	77	14.8	1321	1	US-08-261-	Sequence 3, Applicatio	4.27e+01
19	77	14.8	1321	4	PCT-US95-0	Sequence 3, Applicatio	4.27e+01
20	76	14.6	516	1	US-08-229-	Sequence 2, Applicatio	5.10e+01
21	76	14.6	980	2	US-08-473-	Sequence 6, Applicatio	5.10e+01
22	76	14.6	985	2	US-08-473	Sequence 2, Applicatio	5.10e+01
23	75	14.4	409	2	US-08-185-	Sequence 2, Applicatio	6.08e+01
24	75	14.4	409	1	US-08-191-	Sequence 2, Applicatio	6.08e+01
25	75	14.4	416	1	US-08-117-	Sequence 6, Applicatio	6.08e+01
26	74	14.2	42	1	US-08-118-	Sequence 2[illegible]	7.24e+01
27	74	14.2	42	4	PCT-US93-0	Sequence 2[illegible]	7.24e+01
28	74	14.2	274	2	US-08-758	Sequence 1, Applicati	7.24e+01
29	74	14.2	809	1	US-08-473-	Sequence 2, Applicatio	7.24e+01
30	74	14.2	857	1	US-08-414-	Sequence 1, Applicatio	7.24e+01
31	74	14.2	857	1	US-08-473	Sequence 1, Applicatio	7.24e+01
32	74	14.2	875	1	US-08-220-	Sequence 1, Applicatio	7.24e+01
33	74	14.2	875	1	US-08-220	Sequence 2, Applicatio	7.24e+01
34	74	14.2	879	1	US-08-414-	Sequence 1, Applicatio	7.24e+01
35	74	14.2	879	1	US-08-473-	Sequence 1, Applicatio	7.24e+01
36	74	14.2	879	1	US-08-414-	Sequence 2, Applicatio	7.24e+01
37	74	14.0	311	1	US-08-605-	Sequence 10, Applicati	8.24e+01
38	74	14.0	642	1	US-07-841-	Sequence 4, Applicatio	8.24e+01
39	74	14.0	778	2	US-08-714-	Sequence 1, Applicatio	8.24e+01
40	72	13.8	441	1	US-08-676-	Sequence 4, Applicatio	9.69e+01
41	72	13.8	498	1	US-08-149-	Sequence 5, Applicatio	9.69e+01
42	72	13.8	498	2	US-08-911-	Sequence 5, Applicatio	9.69e+01
43	72	13.8	498	3	US-08-170-	Sequence 5, Applicatio	9.69e+01
44	72	13.8	498	2	US-08-514-	Sequence 8, Applicatio	9.69e+01
45	72	13.8	528	5	5475-45-2	atent No. [illegible]	9.69e+01

ALIGNMENTS

RESULT 1
ID PCT-US93-08528 266 STANDARD; PRT; 42 AA
XX
AC ^^^^^^
XX
DT
XX
DE Sequence 266, Application PC/US9308528
XX
CC Sequence 266, Application PC/US9308528
CC GENERAL INFORMATION:
CC APPLICANT: New York University
CC TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED
CC TITLE OF INVENTION: RECEPTORS, AND COMPOSITIO[illegible]
CC NUMBER OF SEQUENCES: 448
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: BROWDY AND NEIMARK
CC STREET: 419 Seventh Street, N.W., Suite [illegible]
CC CITY: Washington
CC STATE: D.C.
CC COUNTRY: USA
CC ZIP: 20004
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #.
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PC/US93/08528
CC FILING DATE: 09-SEP-1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/943,236
CC FILING DATE: 10-SEP-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Townsend, Kevin G.
CC REGISTRATION NUMBER: 4[illegible]
CC REFERENCE/DOCKET NUMBER: MCBRAY-2 PCT
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202 628 5197
CC TELEFAX: 202 737 3528
CC TELEX: 248633
CC INFORMATION FOR SEQ ID NO: 266:
CC SEQUENCE CHARACTERISTICS
CC LENGTH: 42 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear

```
CC       MOLECULE TYPE:  peptide
SQ   SEQUENCE   42 AA; 4627 MW; 9736 CN;

  Query Match           16.9%;  Score 88;  DB 1;  Length 42;
  Best Local Similarity  37.5%;  Pred. No. 5.82e+00;
  Matches   12;  Conservative    8;  Mismatches  10;  Indels   2;  Gaps   2;

Db       3 IFVLYDFLFCSLGSFYIPIAVILIMLCTASIF 34
           || |  :|| ::|||:|:    ::| |   :||
Qy      10 IF-LNSYLFVPIGSIYVVDIAVFITL-DDAIF 39


RESULT    2
ID   US-08-118-270-266      STANDARD;      PRT;   42 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 266, Application US/08118270
XX
CC    Sequence 266, Application US/08118270
CC    Patent No. 5508384
CC     GENERAL INFORMATION:
CC       APPLICANT:  Murphy, Randall B.
CC       APPLICANT:  Schuster, David I
CC       TITLE OF INVENTION:  POLYPEPTIDES OF G-COUPLED PROTEIN
CC       TITLE OF INVENTION:  RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
CC       NUMBER OF SEQUENCES:  348
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  BROWDY AND NEIMARK
CC         STREET:  419 Seventh Street, N.W., Suite 300
CC         CITY:  Washington
CC         STATE:  D.C.
CC         COUNTRY:  USA
CC         ZIP:  20004
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/118,270
CC         FILING DATE:  09-SEP-1993
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/943,236
CC         FILING DATE:  10-SEP-1992
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Townsend, Kevin G
CC         REGISTRATION NUMBER:  34,033
CC         REFERENCE/DOCKET NUMBER:  MURPHY-2A
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  202-628-5197
CC         TELEFAX:  202-737-3528
CC         TELEX:  248633
CC     INFORMATION FOR SEQ ID NO:  266.
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  42 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  peptide
SQ   SEQUENCE   42 AA; 4627 MW; 9736 CN;

  Query Match           16.9%;  Score 88;  DB 1;  Length 42;
  Best Local Similarity  37.5%;  Pred. No. 5.82e+00;
  Matches   12;  Conservative    8;  Mismatches  10;  Indels   2;  Gaps   2;

Db       3 IFVLYDFLFCSLGSFYIPIAVILIMLCTASIF 34
           || |  :|| ::|||:|:    ::| |   :||
Qy      10 IF-LNSYLFVPIGSIYVVDIAVFITL-DDAIF 39


RESULT    3
ID   US-08-942-001-8      STANDARD;      PRT;   222 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 8, Application US/08942001A
XX
CC    Sequence 8, Application US/08942001A
CC    Patent No. 6020194
CC     GENERAL INFORMATION:
CC      APPLICANT: Mueller, Paul R.
CC      APPLICANT: Coleman, Thomas R.
CC      APPLICANT: Kumagai, Akiko
CC      APPLICANT: Dunphy, William G.
CC      TITLE OF INVENTION: Cdc2 Kinase, Myt 1, and Methods of Use
CC      FILE REFERENCE: 06618/074001
CC      CURRENT APPLICATION NUMBER: US/08/942,001A
CC      CURRENT FILING DATE: 1997-10-01
CC      EARLIER APPLICATION NUMBER: 60/028,073
CC      EARLIER FILING DATE: 1996-10-04
CC      NUMBER OF SEQ ID NOS: 9
CC      SOFTWARE: PatentIn Ver. 2.0
CC     SEQ ID NO 8
CC       LENGTH: 222
CC       TYPE: PRT
CC       ORGANISM: Schizosaccharomyces pombe
SQ   SEQUENCE   222 AA; 25297 MW; 250109 CN;

  Query Match           16.9%;  Score 88;  DB 3;  Length 222;
  Best Local Similarity  28.6%;  Pred. No. 5.82e+00;
  Matches   12;  Conservative   10;  Mismatches  19;  Indels   1;  Gaps   1;

Db     179 LANHLYDKPADIFSLGITVFEAAANIVLPDNGQSWQKLRSGD 220
           | : |:   : |: : |:|| :  : ::|  |  |   | ||
Qy      12 LNSYLFVPIGSIYVVDIAVFITLDDAIFPSIG-CWKVSRKGD 52


RESULT    4
ID   US-08-284-391B-25      STANDARD;      PRT;   182 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 25, Application US/08284391B
XX
CC    Sequence 25, Application US/08284391B
CC    Patent No. 5851828
CC     GENERAL INFORMATION:
CC       APPLICANT:  Seed, Brian
CC       APPLICANT:  Banapour, Babak
CC       APPLICANT:  Romeo, Charles
CC       APPLICANT:  Kolanus, Waldemar
CC       TITLE OF INVENTION:  TARGETED CYTOLYSIS OF HIV-INFECTED
CC       TITLE OF INVENTION:  CELLS BY CHIMERIC CD4 RECEPTOR- BEARING CELLS
CC       NUMBER OF SEQUENCES:  53
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Clark & Elbing LLP
CC         STREET:  176 Federal Street
CC         CITY:  Boston
CC         STATE:  MA
CC         COUNTRY:  USA
CC         ZIP:  02110
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette
CC         COMPUTER:  IBM Compatible
CC         OPERATING SYSTEM:  DOS
CC         SOFTWARE:  FastSEQ for Windows Version 2.0
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/284,391B
CC         FILING DATE:  02-AUG-1994
```

CC CLASSIFICATION: 514
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/195,395
CC FILING DATE: 14-FEB-1994
CC APPLICATION NUMBER: 07/847,566
CC FILING DATE: 06-MAR-1992
CC APPLICATION NUMBER: 07/665,961
CC FILING DATE: 07-MAR-1991
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Elbing, Karen L.
CC REGISTRATION NUMBER: 35,238
CC REFERENCE/DOCKET NUMBER: 00786/247001
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 617-428-0200
CC TELEFAX: 617-428-7045
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 25:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 182 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 182 AA; 20469 MW; 157982 CN;

Query Match 16.7%; Score 87; DB 2; Length 182;
Best Local Similarity 35.7%; Pred. No. 7.00e+00;
Matches 10; Conservative 12; Mismatches 5; Indels 1; Gaps 1;

Db 115 ISGFLFAEIVSIFVLAVGVYFIAGQDGV 142
::::|| | ||:|: :|: || :|::
Qy 12 LNSYLFVPIGSIYVVDIAV-FITLDDAI 38

RESULT 5
ID PCT-US95-00454-25 STANDARD; PRT: 182 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 25, Application PC/TUS9500454
XX
CC Sequence 25, Application PC/TUS9500454
CC GENERAL INFORMATION:
CC APPLICANT: Seed, Brian et al.
CC TITLE OF INVENTION: Targeted Cytolysis of HIV-Infected
CC TITLE OF INVENTION: Cells by Chimeric CD4 Receptor-
CC TITLE OF INVENTION: Bearing Cells
CC NUMBER OF SEQUENCES: 27
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Fish & Richardson
CC STREET: 225 Franklin Street
CC CITY: Boston
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02110-2804
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
CC COMPUTER: IBM PS/2 Model 50Z or 55SX
CC OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
CC SOFTWARE: Wordperfect (Version 5.0)
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/00454
CC FILING DATE:
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 07/847,566
CC FILING DATE: March 6, 1992
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 07/665,961
CC FILING DATE: March 7, 1991
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Clark, Paul T.
CC REGISTRATION NUMBER: 30,162
CC REFERENCE/DOCKET NUMBER: 00786/247001
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 542-5070
CC TELEFAX: (617) 542-8906
CC TELEX: 200154
CC INFORMATION FOR SEQ ID NO: 25:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 182 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: amino acid
SQ SEQUENCE 182 AA; 20469 MW; 157982 CN;

Query Match 16.7%; Score 87; DB 4; Length
Best Local Similarity 35.7%; Pred. No. 7.00e+00;
Matches 10; Conservative 12; Mismatches 5;

Db 115 ISGFLFAEIVSIFVLAVGVYFIAGQDGV 142
::::|| | ||:|: :|: || :|::
Qy 12 LNSYLFVPIGSIYVVDIAV-FITLDDAI 38

RESULT 6
ID 08 08 417 495 25 STANDARD; PRT; 182 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 25, Application US/08417495
XX
CC Sequence 25, Application US/08417495
CC Patent No. 5843728
CC GENERAL INFORMATION:
CC APPLICANT: Seed, Brian et al.
CC TITLE OF INVENTION: Redirection of Cellular
CC TITLE OF INVENTION: Receptor
CC NUMBER OF SEQUENCES: 27
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Fish & Richardson
CC STREET: 225 Franklin Street
CC CITY: Boston
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02110-2804
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
CC COMPUTER: IBM PS/2 Model 50Z or 55SX
CC OPERATING SYSTEM: IBM P.C. DOS (Version 3.
CC SOFTWARE: Wordperfect (Version 5.0)
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/417,495
CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/07/203,866
CC FILING DATE:
CC APPLICATION NUMBER: US/07/847,566
CC FILING DATE:
CC APPLICATION NUMBER: 07/665,961
CC FILING DATE: March 7, 199
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Clark, Paul T.
CC REGISTRATION NUMBER: 30,162
CC REFERENCE/DOCKET NUMBER: 00786/119002
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 542-5070
CC TELEFAX: (617) 542-8906

```
CC          TELEX:  200154
CC      INFORMATION FOR SEQ ID NO:  25:
CC        SEQUENCE CHARACTERISTICS:
CC          LENGTH:  182 amino acids
CC          TYPE:  amino acid
CC          TOPOLOGY:  linear
CC        MOLECULE TYPE:  amino acid
SQ   SEQUENCE   182 AA; 20469 MW; 157982 CN;

  Query Match            16.7%;  Score 87;  DB 2;  Length 182;
  Best Local Similarity  35.7%;  Pred. No. 7.00e+00;
  Matches   10;  Conservative   12;  Mismatches   5;  Indels   1;  Gaps   1;

Db    115 ISGFLFAEIVSIFVLAVGVYFIAGQDGV 142
          :::::||  | ||:|: ::| ||: :|::
Qy     12 LNSYLFVPIGSIYVVDIAV-FITLDDAI 38


RESULT    7
 D   PCT-US92 01785 25     STANDARD;     PRT;   182 AA.
XX
AC   XXXXXX
XX
DT
XX
DE   Sequence 25, Application PC/TUS9201785
XX
CC    Sequence 25, Application PC/TUS9201785
CC     GENERAL INFORMATION:
CC       APPLICANT:  The General Hospital Corporation
CC       TITLE OF INVENTION:  Redirection of Cellular Immunity by Receptor
CC       TITLE OF INVENTION:  Chimeras
CC       NUMBER OF SEQUENCES:  27
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Fish & Richardson
CC         STREET:  225 Franklin Street
CC         CITY:  Boston
CC         STATE:  MA
CC         COUNTRY:  USA
CC         ZIP:  02110-2804
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
CC         COMPUTER:  IBM PS/2 Model 502 or 55SX
CC         OPERATING SYSTEM:  IBM P.C. DOS (Version 3.30)
CC         SOFTWARE:  Wordperfect (Version 5.0)
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/TUS92/01785
CC         FILING DATE:  19920306
CC         CLASSIFICATION:  530
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  07/665,961
CC         FILING DATE:  March 7, 1991
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Clark, Paul T.
CC         REGISTRATION NUMBER:  30,162
CC         REFERENCE/DOCKET NUMBER:  00786/119002
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (617) 542-5070
CC         TELEFAX:  (617) 542-8906
CC         TELEX:  200154
CC     INFORMATION FOR SEQ ID NO:  25:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  182 amino acids
CC         TYPE:  AMINO ACID
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  amino acid
SQ   SEQUENCE   182 AA; 20469 MW; 157982 CN;

  Query Match            16.7%;  Score 87;  DB 4;  Length 182;
  Best Local Similarity  35.7%;  Pred. No. 7.00e+00;
  Matches   10;  Conservative   12;  Mismatches   5;  Indels   1;  Gaps   1;

Db    115 ISGFLFAEIVSIFVLAVGVYFIAGQDGV 142
          :::::||  | ||:|: ::| ||: :|::
Qy     12 LNSYLFVPIGSIYVVDIAV-FITLDDAI 38


RESULT    8
ID   US-08-118-270-268     STANDARD;     PRT;   46 AA.
XX
AC   XXXXXX
XX
DT
XX
DE   Sequence 268, Application US/08118270
XX
CC    Sequence 268, Application US/08118270
CC    Patent No. 5508384
CC     GENERAL INFORMATION:
CC       APPLICANT:  Murphy, Randall B.
CC       APPLICANT:  Schuster, David I.
CC       TITLE OF INVENTION:  POLYPEPTIDES OF G-COUPLED PROTEIN
CC       TITLE OF INVENTION:  RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
CC       NUMBER OF SEQUENCES:  348
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  BROWDY AND NEIMARK
CC         STREET:  419 Seventh Street, N.W., Suite 300
CC         CITY:  Washington
CC         STATE:  D.C.
CC         COUNTRY:  USA
CC         ZIP:  20004
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/118,270
CC         FILING DATE:  09-SEP-1993
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/943,236
CC         FILING DATE:  10-SEP-1992
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Townsend, Kevin G.
CC         REGISTRATION NUMBER:  34,033
CC         REFERENCE/DOCKET NUMBER:  MURPHY-2A
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  202-628-5197
CC         TELEFAX:  202-737-3528
CC         TELEX:  248633
CC     INFORMATION FOR SEQ ID NO:  268:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  46 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  peptide
SQ   SEQUENCE   46 AA; 5091 MW; 13153 CN;

  Query Match            16.1%;  Score 84;  DB 1;  Length 46;
  Best Local Similarity  46.2%;  Pred. No. 1.21e+01;
  Matches   12;  Conservative    9;  Mismatches   3;  Indels   2;  Gaps   2;

Db     19 IFNLMAFLFCSLGSFYI-PIAVILLI 43
          || | ::|| ::||:|:  |||:::|
Qy     10 IF-LNSYLFVPIGSIYVVDIAVFITL 34


RESULT    9
ID   PCT-US93-08528-268     STANDARD;     PRT;   46 AA.
XX
AC   XXXXXX
XX
DT
XX
DE   Sequence 268, Application PC/TUS9308528
```

```
XX
CC    Sequence 268, Application PC/TUS97308528
CC     GENERAL INFORMATION:
CC       APPLICANT:  New York University
CC       TITLE OF INVENTION:  POLYPEPTIDES OF G-COUPLED PROTEIN
CC       TITLE OF INVENTION:  RECEPTORS, AND COMPOSITIONS ANL METHODS THEREOF
CC       NUMBER OF SEQUENCES:  348
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  BROWDY AND NEIMARK
CC         STREET:  419 Seventh Street, N.W., Suite 300
CC         CITY:  Washington
CC         STATE:  D.C.
CC         COUNTRY:  USA
CC         ZIP:  20004
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/US93/08528
CC         FILING DATE:  09-SEP-1993
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/943,236
CC         FILING DATE:  10-SEP-1992
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Townsend, Kevin G.
CC         REGISTRATION NUMBER:  34,033
CC         REFERENCE/DOCKET NUMBER:  MURPHY-2 PCT
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  202-628-5197
CC         TELEFAX:  202-737-3528
CC         TELEX:  248633
CC     INFORMATION FOR SEQ ID NO:  268:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  46 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  peptide
SQ   SEQUENCE   46 AA; 5091 MW; 13151 CN;

  Query Match           16.1%;  Score 84;  DB 4;  Length 46;
  Best Local Similarity  46.2%;  Pred. No. 1.2le-01;
  Matches   12;  Conservative    9;  Mismatches   4;  Indels   2;  Gaps   2;

Db      19 IFNLMAFLFCSLGSFYI-PIAVILTI 43
           || | ::|| ::|:|:  |||::|
Qy      10 IF-LNSYLFVPIGSIYVVDIAVFIIL 34


RESULT   10
ID   US-08-928-692-20      STANDARD;      PRT;   371 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 20, Application US/08928692
XX
CC    Sequence 20, Application US/08928692
CC    Patent No. 5958727
CC     GENERAL INFORMATION:
CC       APPLICANT:  Brody, Howard
CC       APPLICANT:  Yaver, Deborah S.
CC       APPLICANT:  Lamsa, Michael
CC       APPLICANT:  Hansen, Kim
CC       TITLE OF INVENTION:  Methods for Modifying the Production of
CC       TITLE OF INVENTION:  a Polypeptide
CC       NUMBER OF SEQUENCES:  80
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  No. 5958727a No. 5958727disk of No. 5958727th America, Inc.
CC         STREET:  405 Lexington Avenue
CC         CITY:  New York
CC         STATE:  NY
CC         COUNTRY:  USA
CC         ZIP:  10174
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette
CC         COMPUTER:  IBM Compatible
CC         OPERATING SYSTEM:  DOS
CC         SOFTWARE:  FastSEQ for Windows Version 2.0
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/928,692
CC         FILING DATE:  12 SEPT 1997
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Lambiris, Elias J
CC         REGISTRATION NUMBER:  33,728
CC         REFERENCE/DOCKET NUMBER:  4944.200 US
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  212-867-0123
CC         TELEFAX:  212-878-9655
CC     INFORMATION FOR SEQ ID NO:  20:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  371 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  No. 5958727a
SQ   SEQUENCE   371 AA; 41551 MW; 771816 CN;

  Query Match           16.1%;  Score 84;  DB 2;  Length
  Best Local Similarity  46.2%;  Pred. No. 1.2le-01;
  Matches   12;  Conservative    9;  Mismatches   4;  Indels   2;  Gaps   2;

Db     249 ILFLGGYLFINLKDVENILDFRYSNL  FIF-LIN-WMVSFRS
           ::|| :|:: : :: :::: : :  || | |:|: ::
Qy       9 FIFLNSYLFVPIGSIYVVDIAVFIIL-SAIFFSIGYWRVSFRS


RESULT   11
ID   US-08-619-554-2      STANDARD;      PRT;   187 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application US/08619554
XX
CC    Sequence 2, Application US/08619554
CC    Patent No. 5821353
CC     GENERAL INFORMATION:
CC       APPLICANT:  DOUGLAS, Cameron M.
CC       APPLICANT:  CHREBET, Gary L.
CC       APPLICANT:  CLEMAS, Joseph
CC       APPLICANT:  EL-SHERBEINI, Mohammed
CC       APPLICANT:  FOOR, Forrest
CC       APPLICANT:  KAHN, Jennifer,
CC       APPLICANT:  KELLY, Rosemarie,    PARENT, S.A.
CC       APPLICANT:  MARRINAN, Joan,    RAMADAN, N.M.
CC       APPLICANT:  McKIM, Nancy,    SCHLEIER, E.A
CC       APPLICANT:  ONISHI, Janet,    SHEI, Gan-Ju
CC       TITLE OF INVENTION:  DNA ENCODING 1,4 BETA-1,3
CC       TITLE OF INVENTION:  SYNTHASE SUBUNITS
CC       NUMBER OF SEQUENCES:  8
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  JOSEPH A. PIZZA - MERCK & CO.,
CC         STREET:  126 EAST LINCOLN AVENUE - P.O. BOX
CC         CITY:  RAHWAY
CC         STATE:  NJ
CC         COUNTRY:  USA
CC         ZIP:  07065
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette
CC         COMPUTER:  IBM Compatible
```

```
CC         OPERATING SYSTEM:  DOS
CC         SOFTWARE:  FastSEQ for Windows Version 2.0
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/619,554
CC         FILING DATE:  01-AUG-1996
CC         CLASSIFICATION:  536
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:
CC         FILING DATE:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  COPPOLA, JOSEPH A
CC         REGISTRATION NUMBER:  38,413
CC         REFERENCE/DOCKET NUMBER:  19104PI
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  732-594-6734
CC         TELEFAX:  732-594-4720
CC         TELEX:
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1876 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  unknown
CC         TOPOLOGY:  unknown
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   1876 AA; 214849 MW; 18426567 CN;

  Query Match           16.1%;  Score 84;  DB 2;  Length 1876;
  Best Local Similarity  33.3%;  Pred. No. 1.21e+01;
  Matches   12;  Conservative   11;  Mismatches   9;  Indels   4;  Gaps   3;

Db    678 GLVIATDFILFFLDTYLWYIIVNIIFSVGKSFYLGI 713
          || |  |||  ||:::|| :: : :|: |  : :: :
Qy      3 GL-INEDFI- FLNSYL-FVPIGSIYVVDIAVFITL 34


RESULT   12
ID   US-08-619-554-4      STANDARD;      PRT;   1895 AA.
XX
AC   XXXXXX
XX
DT
XX
DE   Sequence 4, Application US/08619554
XX
CC    Sequence 4, Application US/08619554
CC    Patent No. 5821453
CC     GENERAL INFORMATION:
CC       APPLICANT:  DOUGLAS, Cameron M.
CC       APPLICANT:  CHREBET, Gary L.
CC       APPLICANT:  CLEMAS, Joseph
CC       APPLICANT:  EL-SHERBEINI, Mohammed
CC       APPLICANT:  FOOR, Forrest
CC       APPLICANT:  KAHN, Jennifer.
CC       APPLICANT:  KELLY, Rosemarie, - PARENT, S.A.
CC       APPLICANT:  MARRINAN, Jean, - RAMADAN, N.M.
CC       APPLICANT:  MORIN, Nancy, - REGISTER, E.A
CC       APPLICANT:  ONISHI, Janet, - SHEI, Gan-Ju
CC       TITLE OF INVENTION:  DNA ENCODING 1,3 BETA-D GLUCAN
CC       TITLE OF INVENTION:  SYNTHASE SUBUNITS
CC       NUMBER OF SEQUENCES:  8
CC       CORRESPONDENCE ADDRESS.
CC         ADDRESSEE:  JOSEPH A. COPPOLA - MERCK & CO., INC.
CC         STREET:  126 EAST LINCOLN AVENUE   P.O. BOX 2000
CC         CITY:  RAHWAY
CC         STATE:  NJ
CC         COUNTRY:  USA
CC         ZIP:  07065
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette
CC         COMPUTER:  IBM Compatible
CC         OPERATING SYSTEM:  DOS
CC         SOFTWARE:  FastSEQ for Windows Version 2.0
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/619,554
CC         FILING DATE:  01-AUG-1996
CC         CLASSIFICATION:  536
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:
CC         FILING DATE:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  COPPOLA, JOSEPH A
CC         REGISTRATION NUMBER:  38,413
CC         REFERENCE/DOCKET NUMBER:  19104PI
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  732-594-6734
CC         TELEFAX:  732-594-4720
CC         TELEX:
CC     INFORMATION FOR SEQ ID NO:  4:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1895 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  unknown
CC         TOPOLOGY:  unknown
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   1895 AA; 216988 MW; 18982348 CN;

  Query Match           15.9%;  Score 83;  DB 2;  Length 1895;
  Best Local Similarity  26.5%;  Pred. No. 1.45e+01;
  Matches   9;  Conservative   14;  Mismatches   9;  Indels   3;  Gaps   2;

Db    699 MIATDFILFFLDTYLWYIVVNTVFSVGKSFYLGI 732
          :|  |||  ||:::|| :: : ::: |  : :: :
Qy      4 LINEDFI--FLNSYL-FVPIGSIYVVDIAVFITL 34


RESULT   13
ID   US-07-857-224B-38      STANDARD;      PRT;   269 AA.
XX
AC   XXXXXX
XX
DT
XX
DE   Sequence 38, Application US/07857224B
XX
CC    Sequence 38, Application US/07857224B
CC    Patent No. 5958784
CC     GENERAL INFORMATION:
CC       APPLICANT:  Benner, Steven A.
CC       TITLE OF INVENTION:  Predicting Folded Structures of Proteins
CC       NUMBER OF SEQUENCES:  114
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Steven A. Benner
CC         STREET:  Hadlaubstrasse 151
CC         CITY:  Zurich
CC         STATE:  none
CC         COUNTRY:  Switzerland
CC         ZIP:  (note: this is an international post code) CH 8092
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  3.5 inch diskette, 1.4 Mb storage
CC         COMPUTER:  Apple MacIntosh
CC         OPERATING SYSTEM:  MacIntosh 7.0
CC         SOFTWARE:  Microsoft Word
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/07/857,224B
CC         FILING DATE:  03/25/92
CC         CLASSIFICATION:  436
CC       PRIOR APPLICATION DATA:  none
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (International) 41 1 632 2830
CC         TELEFAX:  (International) 41 1 262 2437
CC         TELEX:  none
CC     INFORMATION FOR SEQ ID NO:  38:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  269
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
```

```
CC       MOLECULE TYPE:
CC         DESCRIPTION:  protein
CC       ORIGINAL SOURCE:
CC         ORGANISM:  human
CC       FEATURE:  Protein kinase; Table 8 Column 43
CC       PUBLICATION INFORMATION:
CC         AUTHORS:
CC         AUTHORS:  Banks, S. K.
CC         AUTHORS:  Quinn, A. M.
CC         AUTHORS:  Hunter, T.
CC         TITLE:  The protein kinase family
CC         JOURNAL:  Science
CC         VOLUME:  241
CC         PAGES:  42-52
CC         DATE:  1988
SQ   SEQUENCE   269 AA; 30511 MW; 869255 CN;

  Query Match           15.5%;  Score 81;  DB 2;  Length 269;
  Best Local Similarity  28.6%;  Pred. No. 2.0e-01;
  Matches   12;  Conservative    9;  Mismatches  20;  Indels   1;  Gaps   1;

Db     175 LANHLYDKPADIWSLGITVFEAAANIVLPDNGQSWQKLRSGD 216
           |  :|:   :|  :|:||  :  ::|  |  |   | ||
Qy      12 LNSYLFVPIGSIYVVDIAVFILDDAIPPSIG-CWKVSRKGD 52


RESULT   14
ID   US-08-385-590A-18      STANDARD;      PRT;   69 AA.
XX
AC   xxxxxx
XX
D1
XX
DE   Sequence 18, Application US/08385590A
XX
CC    Sequence 18, Application US/08385590A
CC    Patent No. 5824861
CC     GENERAL INFORMATION:
CC       APPLICANT:  Aldwinckle, Herbert S.
CC       APPLICANT:  No. 5824861-Elli, John L.
CC       TITLE OF INVENTION:  TRANSGENIC POMACEOUS FRUIT WITH FIRE
CC       TITLE OF INVENTION:  BLIGHT RESISTANCE
CC       NUMBER OF SEQUENCES:  28
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Nixon, Hargrave, Devans & Doyle LLP
CC         STREET:  Clinton Square
CC         CITY:  Rochester
CC         STATE:  New York
CC         COUNTRY:  U.S.A.
CC         ZIP:  14603
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  Patentin Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/385,590A
CC         FILING DATE:  08-FEB-1995
CC         CLASSIFICATION:  800
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/954,342
CC         FILING DATE:  30-SEP-1992
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Goldman, Michael L.
CC         REGISTRATION NUMBER:  30,727
CC         REFERENCE/DOCKET NUMBER:  19603/143
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (716) 263-1304
CC         TELEFAX:  (716) 263-1600
CC     INFORMATION FOR SEQ ID NO:  18:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  69 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  unknown
CC       MOLECULE TYPE:  peptide
SQ   SEQUENCE   69 AA; 7680 MW; 21A7F CN;

  Query Match           [illegible]  Score 79;  DB 2;  Length
  Best Local Similarity  21.4%;  Pred. No. 2.9e-01;
  Matches    9;  Conservative    [illegible]  Mismatches  [illegible]

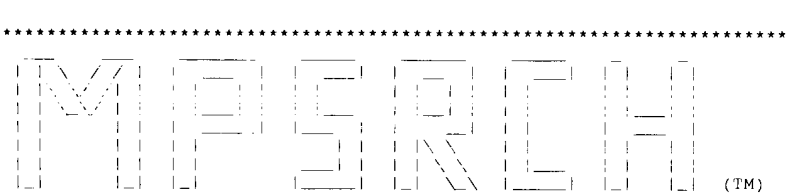
Db       4 FLFSEMPSFPLVSTLLLFLIISHSSAATMPRWRLFRRIDRVS 45
           :||  : |:::|  : :||: :  :  :::  |:: |: |: :
Qy      15 YIFVPIGSIYVVDIAVFILDDAIPPSIGCWKVSRKGDKSS 55


RESULT   15
ID   US-08-547-182-2      STANDARD;      PRT;   418 AA.
XX
AC   xxxxxx
XX
D1
XX
DE   Sequence 2, Application US/08547182
XX
CC    Sequence 2, Application US/08547182
CC    Patent No. 5672694
CC     GENERAL INFORMATION:
CC       APPLICANT:  Campbell, Kevin P.
CC       APPLICANT:  Lim, Leonard
CC       APPLICANT:  Duclos, Franck
CC       APPLICANT:  Sunada, Yoshitaka
CC       APPLICANT:  Bockmann, Jacques S.
CC       APPLICANT:  Brown, Odile
CC       APPLICANT:  Tome, Fernando M.S.
CC       APPLICANT:  Fardeau, Marcel
CC       APPLICANT:  Jackson, Charles E.
CC       TITLE OF INVENTION:  CALPAIN-LIKE NUCLEIC ACID
CC       TITLE OF INVENTION:  DISTINCT MUTATIONS AND
CC       NUMBER OF SEQUENCES:  12
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Kevin M. Farrell
CC         STREET:  P.O. Box 999
CC         CITY:  York Harbor
CC         STATE:  ME
CC         COUNTRY:  USA
CC         ZIP:  03911
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  Patentin Release #1.0, Version #1
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/547,182
CC         FILING DATE:
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Farrell, Kevin M.
CC         REGISTRATION NUMBER:  35,555
CC         REFERENCE/DOCKET NUMBER:  PIRF-9502
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  207-363-0558
CC         TELEFAX:  207-363-0528
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  318 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   318 AA; 34777 MW; 732444 CN;

  Query Match           15.1%;  Score 79;  DB 1;  Length
  Best Local Similarity  31.3%;  Pred. No. 2.9e-01;
  Matches   15;  Conservative   16;  Mismatches  26;
```

```
Db      65 AICVI-ILLFILAVINLIITLVIWAVIRIGPNGCD-SMEFHESGLLRF 110
           :| |: | :||     :: ::  | | | | :| : |::|| ||:::
Qy      22 SIYVVDIAVFITLDDAIEPSTGCWKVSRKGDKGGECSLDFH-CGLFQL 68

Search completed: Fri Aug 25 17:59:25 2000
Job time : 20 secs.
```



Release 4.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993 1998 University of Edinburgh, U K
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 25 17:55:51 2000; MasPar time 3.96 Seconds
406.340 Million cell updates/sec

Tabular output not generated.

Title: >US-09-257-585-5
Description: (1-68) from US09257585.pep
Perfect Score: 522
Sequence: 1 MAGLINEDFIFLNSYLFVPI.....RKGDKGGECSIDFHQGIFQL 68

Scoring table: PAM 150
Gap 11

Searched: 188963 seqs, 24686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq39
1:geneseqp

Statistics: Mean 27.436; Variance 111.070; scale 0.247

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	99	19.0	873	1	R79459	Infectious laryngotrac	1.29e+00
2	94	18.0	825	1	R22237	Sequence of the "gB" h	3.33e+00
3	92	17.6	1088	1	R88635	Plasmid pASK75 open re	4.85e+00
4	92	17.6	1221	1	R52699	Sequence translated fr	4.85e+00
5	92	17.6	1277	1	R52700	Plasmid pASK60 Strep r	4.85e+00
6	91	17.4	527	1	W98268	H. pylori CHP3 33% pro	5.85e+00
7	88	16.9	42	1	W02917	G-protein coupled rece	1.02e+01
8	88	16.9	42	1	R50725	G-protein coupled rece	1.02e+01
9	87	16.7	182	1	W02217	T cell receptor T3 gam	1.23e+01
10	87	16.7	182	1	R27280	T cell receptor protei	1.23e+01
11	87	16.7	182	1	R89447	T3 gamma receptor	1.23e+01
12	87	16.7	182	1	R78670	T3 gamma receptor prot	1.23e+01
13	87	16.7	182	1	W83144	T3 gamma receptor prot	1.23e+01
14	84	16.1	46	1	R50727	G-protein coupled rece	2.13e+01
15	84	16.1	46	1	W02919	G-protein coupled rece	2.13e+01
16	84	16.1	430	1	W88344	Salmonella enterica O	2.13e+01
17	84	16.1	1876	1	R72853	S. cerevisiae FKS .	2.14e+01
18	83	15.9	69	1	R53356	sPR1-Shiva 1 fusion pr	2.55e+01
19	83	15.9	1084	1	W33818	Arabidopsis cellulose	2.55e+01
20	83	15.9	1895	1	R72854	S. cerevisiae FKS2.	2.55e+01
21	81	15.5	116	1	Y05503	Bacillus subtilis meta	3.66e+01
22	81	15.5	211	1	Y05502	Bacillus subtilis meta	3.66e+01
23	81	15.5	312	1	W60095	Synechocystis sp. beta	3.66e+01
24	81	15.5	426	1	Y05504	Bacillus sub	[illegible]
25	79	15.1	235	1	W62549	Soluble lep	[illegible]
26	79	15.1	318	1	W32064	Human beta	[illegible]
27	79	15.1	318	1	W32065	H. sapiens	[illegible]
28	79	15.1	318	1	W 2065	Human beta	[illegible]
29	79	15.1	318	1	W 2065	Human beta	[illegible]
30	79	15.1	318	1	W47382	Beta-Sarcog	[illegible]
31	78	14.9	68	1	R53452	sPR1-Shiva 1	[illegible]
32	78	14.9	357	1	W14774	Nitrile hyd	[illegible]
33	78	14.9	366	1	R12466	HLA-C exon	[illegible]
34	77	14.8	259	1	W20187	H. pylori a	[illegible]
35	77	14.8	426	1	W20654	H. pylori s	[illegible]
36	77	14.8	594	1	R92279	Human kidne	[illegible]
37	77	14.8	1294	1	W75798	Arabidopsis	[illegible]
38	77	14.8	1321	1	R98429	Arabidopsis	[illegible]
39	76	14.6	134	1	W93604	Tomato Xa21	[illegible]
40	76	14.6	366	1	Y07044	Breast cance	[illegible]
41	76	14.6	366	1	R12466	HLA-C exon	[illegible]
42	76	14.6	518	1	R98894	Hydrogen pe	[illegible]
43	76	14.6	788	1	W97419	A membrane	[illegible]
44	76	14.6	908	1	W30611	Arabidopsis	[illegible]
45	76	14.6	985	1	W30607	Arabidopsis	[illegible]

ALIGNMENTS

RESULT 1
ID R79459 standard; Protein; 873 AA.
AC R79459;
DT 16-JAN-1996 (first entry)
DE Infectious laryngotracheitis virus gB glycoprotein.
KW Infectious laryngotracheitis virus; gB; glycoprotein;
KW vaccine; immunoassay; detection; identification; a
KW fowl; probe; antisense; gene expression.
OS Gallid Herpesvirus 1.
FH Key Location/Qualifiers
FT misc_difference 656..659
FT /note= "these amino acids are miss
FT decoded; polypeptide sequence is co
FT specification."
PN US5443831-A.
PD 22-AUG-1995.
PF 29-OCT-1991; 788123.
PR 29-OCT-1991; US 788123.
PR 19-NOV-1993; US-156866
PA (OYDE) UNIV DELAWARE.
PI Keeler CL, Poulsen DJ;
DR WPI; 95-302091/39.
DR N-PSDB; Q97351.
PT Isolated infectious laryngotracheitis virus gB nucl
PT develop prods. for detection and for vaccine(s). pa
PT immunising fowl.
PS Claim 1; Columns 11-16; 11pp; English.
CC The infectious laryngotracheitis virus (ILV) gB gl
CC nucleic acids can be used as probes, for antisense
CC gene expression or for production of gB polypeptide
CC polypeptides can be used for the production of ant
CC used in immunoassays and in vaccines. Recombinant
CC expressing the ILV gB can be used for immunising f
CC chickens.
SQ Sequence 873 AA;

Query Match 19.0%; Score 99; DB 1; Lengt
Best Local Similarity 24.4%; Pred. No. 1.29e+
Matches 9; Conservative 1; Mismatches 11; I

Db 607 LKGENYLKYEDYTPVEVGVNAGIIEVGVSPINLNAIIL 64
Qy 4 LINEDFIFLNSYLFV 18

RESULT 2
ID R22237 standard; Protein; 825 AA.

```
AC   R22237;
DT   29-JUL-1992  (first entry)
DE   Sequence of the "gB" homologue of ILTV encoded by the EcoRI "U" (unique)
DE   genome fragment
KW   Subunit vaccine; immunogen; glycoprotein; promoter.
OS   Infectious laryngotracheitis virus.
FH   Key             Location/Qualifiers
FT   peptide         1. .16
FT                   /label= signal
PN   WO9203554-A.
PD   05-MAR-1992.
PF   28-AUG-1991; AU0383.
PR   24-AUG-1990; AU-001937.
PA   (WEBS-) WEBSTER A PTY LTD.
PI   Sheppard MC, Prideaux C, Johnson M, Fahey KJ, York JJ,
PI   Kongsuwan K;
DR   WPI; 92-096898/12.
DR   N-PSDB; Q22989.
PT   Vaccines against ILTV for use in chickens - comprises
PT   non-infectious sub-units or live recombinant viruses and may be
PT   administered by aerosol
PS   Example; Fig 11; 122pp; English.
CC   The inventors claim a non-infectious subunit vaccine for use against
CC   ILTV which comprises a glycoprotein of ILTV. They also claim DNA
CC   encoding the vaccine, a synthetic polypeptide displaying the
CC   antigenicity of the 205K complex or 60K ILTV glycoprotein and a
CC   recombinant ILTV with heterologous DNA inserted into a non essential
CC   region of the genome; and a recombinant DNA mol. comprising an ILTV
CC   promoter region operatively linked to a heterologous DNA sequence.
CC   The promoter region is the ILTV gp60 promoter, gp205 (gB) promoter
CC   or the OEF3 promoter region. EcoRI-BFA is located 5' of the gB
CC   gene.
SQ   Sequence   825 AA;

  Query Match           18.0%;  Score 94;  DB 1;  Length 825;
  Best Local Similarity  24.4%;  Pred. No. 3.33e+00;
  Matches    9;  Conservative   15;  Mismatches   12;  Indels    1;  Gaps    1;

Db    559 LLGENYLLYEDYTFVRQVNASETEEVSTFINLNATIL 595
          |: |:::: : | ||   : :   : ::: || |: :|:
Qy      4 LINEDFIFLNSYLFV-PIGSIYVVDIAVFTTLDDAIF 39


RESULT    3
ID   R88635 standard; Protein; 1088 AA
AC   R88635;
DT   27 JUN 1996  (first entry)
DE   Plasmid pASK75 open reading frame (b) translation.
KW   Tetracycline; resistance; TetR gene; repressor protein; TetA gene;
KW   prokaryotic expression vector, beta lactamase promoter; bla gene;
KW   circular; cyclic.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_difference 1. .1088
FT                   /note= "Xaa X corresponds to a termination codon"
FT   peptide         36. .42
FT                   /label= TetA
PN   WO9532295 A1.
PD   30-NOV-1995.
PF   17-MAY-1995; E01862.
PR   19-MAY-1994; DE-417598.
PA   (PLAC ) MAX PLANCK GES F FORDERUNG WISSENSCHAFTEN.
PI   Skerra A, Wardenberg C;
DR   WPI; 96-020588/02.
DR   N-PSDB; T11192.
PT   Prokaryotic vector for regulated prodn. of heterologous protein gene
PT   - controlled by promoter repressed by tetracycline repressor
PT   protein, prodn. of which is controlled by independent promoter not
PT   regulated by this repressor
PS   Disclosure; Fig 1a; 50pp; German.
CC   Plasmid pASK75 is a specifically claimed example of a prokaryotic
CC   expression vector comprising the tetracycline promoter/operator (P/O)
CC   region and the tetracycline repressor (tetR) gene. The tetR gene is
CC   under the control of the beta lactamase (bla) promoter. Expression of
CC   heterologous genes inserted downstream of the tet P/O is controlled
CC   by the activity of the Tet repressor. Plasmid pASK75 was derived
CC   from pASK60-Strep (Schmidt and Skerra, Protein. Eng. 6 (1993), 109-
CC   122) and further contains the OmpA leader peptide coding sequence, a
CC   polylinker and a streptavidin-binding peptide strep tag coding
CC   region from the parent plasmid. In addition, pASK75 contains the
CC   ColE1 origin of replication, the bla gene, intergene regions of
CC   filamentous phage f1 and a lipoprotein terminator.
CC   The specification includes the translated sequences from all three
CC   reading frames of pASK75; the TetA coding region is in reading
CC   frame (b).
SQ   Sequence   1088 AA;

  Query Match           17.6%;  Score 92;  DB 1;  Length 1088;
  Best Local Similarity  21.1%;  Pred. No. 4.85e+00;
  Matches    8;  Conservative   19;  Mismatches   10;  Indels    1;  Gaps    1;

Db    295 FIFLNTFKYVSAHETITLINASIILKKEEYEYSTFPCR 332
          |||||:: :|:   :| ::: :::  ::  :::: |:
Qy      9 FIFLNSYLFVPI-GSIYVVDIAVFTTLDDAIFPSIGCW 45


RESULT    4
ID   R52698 standard; Protein; 1221 AA.
AC   R52698;
DT   11 JAN 1995  (first entry)
DE   Sequence translated from reading frame b of plasmid pASK46.
KW   Streptavidin binding peptide; fusion protein; immunoglobulin; Ig;
KW   D1.3Fv; heavy chain; light chain; variable region;
KW   affinity chromatography, purification, peptide tag, detection.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   region          1. .1221
FT                   /note= "translated from reading frame b of pASK46;
FT                   all X's correspond to nonsense codons. The
FT                   specification includes translations of all
FT                   3 reading frames but only b is relevant"
FT   protein         65. .228
FT                   /label= VH_chain
FT   protein         231. .387
FT                   /label= VL_chain
PN   GB2272698-A.
PD   25 MAY-1994.
PF   01-NOV-1993; 022501.
PR   03 NOV 1992; DE 237113.
PA   (KUFN-) KUFNER NONSTOKUKTION GMBH & CO KG STAUS.
PA   (BIOA-) INST BIOANALYTIK GEMEINNUETZIGE GMBH.
PI   Schmidt T, Skerra A;
DR   WPI; 94-153484/19.
DR   N-PSDB; Q62675.
PT   New fusion peptide(s) - have easily controlled binding properties
PT   and are capable of binding to streptavidin
PS   Disclosure; Fig 1; 53pp; English.
CC   Six derivatives of pASK46 (a plasmid for expression of the D1.3Fv
CC   fragment in E.coli) were produced which encode 4 different peptides
CC   at the C terminus of VH and 2 peptides at the C-terminus of VL of
CC   the D1.3Fv fragment. The peptides fused to the C-terminus are all
CC   examples of streptavidin-binding peptides corresponding to a generic
CC   formula (see R52698). The peptides do not interfere with the
CC   protein function but facilitate purification by conferring
CC   streptavidin binding properties on the fusion protein.
SQ   Sequence   1221 AA;

  Query Match           17.6%;  Score 92;  DB 1;  Length 1221;
  Best Local Similarity  21.1%;  Pred. No. 4.85e+00;
  Matches    8;  Conservative   19;  Mismatches   10;  Indels    1;  Gaps    1;

Db    592 FIFLNTFKYVSAHETITLINASIILKKEEYEYSTFPCR 629
          |||||:: :|:   :| ::: :::  ::  :::: |:
Qy      9 FIFLNSYLFVPI-GSIYVVDIAVFTTLDDAIFPSIGCW 45
```

```
RESULT    5
ID   R52700 standard; Protein; 1277 AA.
AC   R52700;
DT   11-JAN-1995  (first entry)
DE   Plasmid pASK60-Strep reading frame "a" translation.
KW   Streptavidin binding peptide; fusion protein; pASK60 Strep;
KW   affinity chromatography; purification; peptide tag; detection.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   region          1..1277
FT                   /note= "translated from reading frame 'a' of
FT                   pASK60-Strep; all X's correspond to nonsense
FT                   codons"
FT   peptide         430..486
FT                   /label= LacZ
PN   GB2272698-A.
PD   25-MAY-1994.
PF   01-NOV-1993; 9322501.
PR   03-NOV-1992; DE-4237113.
PA   (KUEH-) KUEHN KONSTRUKTION GMBH & CO KG KLAUS.
PA   (BIOA-) INST BIOANALYTIK GEMEINNUETZIGE GMBH.
PI   Schmidt I;  Skerra A;
DR   WPI; 94-153484/19.
DR   N-PSDB; Q62676.
PT   New fusion peptide(s) - have easily controlled binding properties
PT   and are capable of binding to streptavidin
PS   Example 5; Fig 8; 53pp; English.
CC   Plasmid pASK60-Strep was produced starting from pASK40
CC   (Biotechnology 9, 273-278, 1991) using site-directed mutagenesis
CC   and PCR. The plasmid contains an improved set of unique restriction
CC   sites, including two sites located directly at the 3' end of the
CC   region coding for the OmpA signal peptide. The polylinker is
CC   followed by a DNA sequence coding for a streptavidin binding
CC   peptide. The plasmid pASK60-Strep is useful for the expression of
CC   polypeptides fused to a streptavidin binding peptide. The
CC   production of such fusion proteins can be specifically detected
CC   using a streptavidin-alkaline phosphatase conjugate.
SQ   Sequence   1277 AA;

  Query Match             17.6%;  Score 92;  DB 1;  Length 1277;
  Best Local Similarity   21.1%;  Pred. No. 4.85e+00;
  Matches    8;  Conservative   19;  Mismatches  10;  Indels   1;  Gaps   1;

Db    689 FIFLNTFKYVSAHETITLINASIILKKEEYEYSTFPCR 726
          ||||||:: :|:   :| ::: :::: :: :::: |:
Qy      9 FIFLNSYLFVPI-GSIYVVDIAVFILDDAIFPSIGCW 45


RESULT    6
ID   W98268 standard; Protein; 527 AA.
AC   W98268;
DT   31-MAR-1999  (first entry)
DE   H. pylori GHPO 335 protein.
KW   GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW   peptic ulcer disease.
OS   Helicobacter pylori.
PN   WO9843478-A1.
PD   08-OCT-1998.
PF   01-APR-1998; 98WO-US006371.
PR   29-JUL-1997; 97US-0902615.
PR   01-APR-1997; 97US-0833457.
PR   24-JUN-1997; 97US-0881227.
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (INMR ) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
PI   Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
DR   WPI; 98-542293/46.
DR   N-PSDB; X13987.
PT   New isolated Helicobacter polynucleotides - used to develop products
PT   for the diagnosis, prevention and treatment of Helicobacter
PT   infections and gastrointestinal diseases
PS   Claim 8; Page 268-271; 2054pp; English.
CC   This sequence represents a Helicobacter pylori GHPO protein of the
CC   invention. The polypeptides can be used for preventing or treating
CC   Helicobacter infections, and gastroduodenal diseas-
CC   these infections, including acute, chronic, and at-
CC   peptic ulcer diseases, e.g. gastric and duodenal u-
CC   used for the production of antibodies. The product-
CC   detection and diagnosis.
SQ   Sequence   527 AA;

  Query Match             17.4%;  Score 91;  DB 1;  Leng-
  Best Local Similarity   31.9%;  Pred. No. 5.85e+00;
  Matches   12;  Conservative    9;  Mismatches  14;

Db    329 FLKEYLYIPLGGNRVKELIVYPNI-LVFLIGGFW 362
          ||: |:::::: :| :: |:     :: : |
Qy     11 FLNSYLFVPIGSIYVVDIAVFIIIIAIFPSIGCW 45


RESULT    7
ID   W02917 standard; peptide; 42 AA.
AC   W02917;
DT   04-OCT-1996  (first entry)
DE   G-protein coupled receptor TM5 consensus polypepti-
KW   G-protein coupled receptor; ligand binding assay;
KW   schizophrenia; dopamine; cAMP; adenosine; thrombin;
KW   muscarinic acetylcholine; endothelin; bombesin; en-
KW   odorant; cytomegalovirus; serotonin.
OS   Synthetic.
PN   US5508384-A.
PD   16-APR-1996.
PF   10-SEP-1992; 944236.
PR   10-SEP-1992; US-944236.
PR   09-SEP-1994; US-11827[illegible].
PA   (UYNY ) UNIV NEW YORK STATE.
PI   Murphy RB,  Schuster DI;
DR   WPI; 96-208785/21.
PT   New dopamine receptor peptide - useful as antipsyc-
PT   for treating schizophrenia
PS   Disclosure; Column 277-280; 184pp; English.
CC   Polypeptides W02747-W02910 are based on the consen-
CC   domain III sequence from G-protein coupled recepto-
CC   and can be used in GPR ligand binding assays. The
CC   to identify fragments, pref. transmembrane fragmen-
CC   (see W02657-W02635) which retain biological activi-
CC   GPR ligand or modulating a GPR ligand binding to a
CC   for examples of polypeptide fragments). The polype-
CC   used in compsns. for treating subjects suffering fr-
CC   to a GPR abnormality e.g. schizophrenia.
SQ   Sequence   42 AA;

  Query Match             17.9%;  Score 88;  DB 1;  Leng-
  Best Local Similarity   47.5%;  Pred. No. 1.02e+01;
  Matches   12;  Conservative    8;  Mismatches  10;

Db      3 IFVLYDFLFCSLGSFYIPIAVILIMLYIASIF 34
          || |  :|| ::|||:|:    ::| |   :||
Qy     10 IF-LNSYLFVPIGSIYVVDIAVFIILDDAIF 39


RESULT    8
ID   R50725 standard; peptide; 42 AA.
AC   R50725;
DT   13-MAY-1996  (first entry)
DE   G-protein coupled receptor TM5 consensus polypepti-
KW   G-protein coupled receptor; ligand binding assay;
KW   psychotic disorder; schizophrenia; dopamine; cAMP;
KW   muscarinic acetylcholine; adrenergic; endothelin;
KW   rhodopsin; opsin; odorant; cytomegalovirus; seroto-
OS   Synthetic.
PN   WO9405695-A1.
PD   17-MAR-1994.
PF   09-SEP-1993; 008528.
PR   10-SEP-1992; US-944236.
PA   (UYNY ) UNIV NEW YORK STATE.
PI   Murphy RB,  Schuster DI;
```

```
DR   WPI; 94-101120/12
FT   Polypeptides of G-coupled receptor proteins (GPRs) - useful for
PT   binding GPR ligands or modulating GPR binding
PS   Claim 9; Page 50; 166pp; English.
CC   Polypeptides R50719-83 are based on the consensus transmembrane domain V
CC   sequence from G-protein coupled receptor (GPR) proteins, and can be used
CC   in G-protein coupled receptor ligand binding assays. The assay can be
CC   used to identify fragments pref. transmembrane fragments, from GPR
CC   proteins (see R48588-R48758 for examples) which retain biological
CC   activity such as binding a GPR ligand or modulating GPR ligand binding to
CC   a GPR (see R48759-R48758, R50569-R50807 and R89189-R89195 for examples of
CC   polypeptide fragments). The polypeptide fragments can be used in
CC   compositions for treating subjects suffering from a pathology related to
CC   a GPR abnormality e.g. a psychotic disorder such as schizophrenia.
SQ   Sequence   42 AA;

  Query Match           16.9%;  Score 88;  DB 1;  Length 42;
  Best Local Similarity  37.5%;  Pred. No. 1.02e+01;
  Matches   12;  Conservative    9;  Mismatches  10;  Indels   2;  Gaps   2;

Db       3 IFVLYDFLFCSLGSFYIPIAVILIMLCTASIF 34
           || |  :|| ::|||:|:    ::| |   :||
Qy      10 IF-LNSYLFVPIGSIYVVDIAVFITL-DDAIF 39


RESULT    9
ID   W02217 standard; Protein; 182 AA.
AC   W02217;
DT   11-NOV-1996  (first entry)
DE   T cell receptor T3 gamma protein.
KW   Chimaeric Receptor, Cellular immunity, adoptive immunotherapy, CD4,
KW   human immunodeficiency virus type 1, HIV-1, AIDS, therapy,
KW   T-cell receptor; T3 gamma, cytotoxic T lymphocyte; CTL.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   domain          141. .182
FT                   /label= Cytolytic_signal_transduction_domain
PN   WO9625953-A1.
PD   29-AUG-1996
PF   25-JAN-1996; 96US-000856
PR   24-FEB-1995; US-394178.
PA   (GEHO ) GEN HOSPITAL CORP.
PI   Kolanus W,  Romeo C,  Seed B;
DR   WPI; 96-402134/40.
PT   Direction of cellular immune response using therapeutic cell
PT   expressing 2 chimaeric receptors - comprising region binding to
PT   target cell and region that signals target cell destruction, or CD28
PT   region, partic. for eliminating HIV-infected cells
PS   Example 10; Fig 17; 120pp; English
CC   Novel chimaeric receptors may contain intracellular and
CC   transmembrane signal transducing domains derived from the T cell
CC   receptor proteins CD3 delta (W02216) or T3 gamma (W02217), or the
CC   B cell receptor proteins mb1 (W02218) or B29 (W02219), and an
CC   extracellular domain that allows specific recognition of, and
CC   binding to, a target cell or target infectious agent. Chimaeric
CC   receptors (see also W02213-15) can be used to redirect the
CC   cellular immunity of a mammal.
SQ   Sequence   182 AA;

  Query Match           16.7%;  Score 87;  DB 1;  Length 182;
  Best Local Similarity  35.7%;  Pred. No. 1.23e+01;
  Matches   10;  Conservative   12;  Mismatches   5;  Indels   1;  Gaps   1;

Db     115 ISGFLFAEIVSIFVLAVGVYFIACQDGV 142
           :...||  | || |:  :: | ||   : :
Qy      12 LNSYLFVPIGSIYVVDIAV-FITLDDAI 38


RESULT   10
ID   R27280 standard; Protein; 182 AA.
AC   R27280;
DT   31-JUL-1995  (first entry)
DE   T cell receptor protein, T3 gamma

KW   Fusion protein; CD4; extracellular domain; zeta, eta, gamma; primer;
KW   polymerase chain reaction; PCR; amplify; human; IgG1; receptor; chimera;
KW   heavy; light; chain; tonsil; promoter; bicistronic mRNA; grp78;
KW   5' untranslated region; 78 kD glucose related protein; COS cell;
KW   immune system; pathogen; tumour cell; HIV; cytotoxic T lymphocyte.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   peptide          141. .182
FT                   /note= "Preferred cytolytic signal transducing
FT                   portion"
PN   WO9215322-A.
PD   17-SEP-1992.
PF   06-MAR-1992; 92US-001785.
PP   07-MAR-1991; US-665961.
PR   06-MAR-1992; WO-US01785.
PA   (GEHO ) GEN HOSPITAL CORP.
PI   Kolanus W,  Romeo C,  Seed B;
DR   WPI; 92-331474/40.
PT   Therapeutic cells expressing chimeric receptors - directing
PT   cellular response to an infective agent, useful in treating
PT   HIV-1, AIDS Pneumocystis carinii infections etc.
PS   Example 10; Fig 17; 114pp; English.
CC   The sequences given in R27279-82 represent the intracellular and
CC   transmembrane signal transducing domains derived form the T cell
CC   receptor protein CD3 delta and T3 gamma, and the B cell receptor
CC   proteins, mb1 and B29. Chimeric receptor proteins may be prepared
CC   by the method of the invention which contain at least the portion of
CC   the sequence which is sufficient for cytolytic signal transduction.
CC   Immune system cells "armed" with a plasmid encoding a chimera such
CC   as this, would respond to the presence of the pathogen appropriate
CC   to their lineage or with tumour cells and immune response could be
CC   beneficially elevated. This method may be used to direct cellular
CC   response to an HIV infected cell by administering to a patient an
CC   effective amount of cytotoxic T lymphocytes which are capable of
CC   specifically recognising and lysing cells infected with HIV, and can
CC   therefore be used in the treatment of HIV infected individuals.
SQ   Sequence   182 AA,

  Query Match           16.7%;  Score 87;  DB 1;  Length 182;
  Best Local Similarity  35.7%;  Pred. No. 1.23e+01;
  Matches   10;  Conservative   12;  Mismatches   5;  Indels   1;  Gaps   1;

Db     115 ISGFLFAEIVSIFVLAVGVYFIACQDGV 142
           :...||  | || |:  :: | ||   : :
Qy      12 LNSYLFVPIGSIYVVDIAV-FITLDDAI 38


RESULT   11
ID   R89447 standard; Protein; 182 AA.
AC   R89447;
DT   26-SEP-1996  (first entry)
DE   T3 gamma receptor.
KW   CD7, transmembrane domain, chimeric receptor, CD5, CD34, CH2; CH3; IgG1;
KW   human; CD4; HIV; proteinaceous alpha helix, T cell, B cell; neutrophil;
KW   dendritic cell; therapy; mammal; infection; CD3 delta; T3 gamma; mb1;
KW   B29.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   domain          141. .182
FT                   /note= "cytolytic signal transducing domain"
PN   WO9603883-A1.
PD   15-FEB-1996.
PF   26-JUL-1995; 95US-009468.
PR   02-AUG-1994; US-284391.
PP   24-FEB-1995; US-394388.
PA   (GEHO ) GEN HOSPITAL CORP.
PI   Banapour B,  Kolanus W,  Romeo C,  Seed B;
DR   WPI; 96-129034/13.
PT   Membrane bound chimeric receptor comprising extracellular portion
PT   including CD4 fragment - cells expressing receptor can be used for
PT   treatment of HIV infection
PS   Example 12; Fig 17; 134pp; English.
CC   R89446-R89449 represent intracellular and transmembrane signal
```

CC transducing domains that can be used in the membrane bound proteinaceous
CC chimeric receptor of the invention. This sequence represents the T3
CC gamma receptor protein. Alternatively the transmembrane region of the
CC chimeric receptor contains a portion of the CD7, CD5 or CD34
CC transmembrane domains. The extracellular portion of the receptor can
CC also be separated from the intracellular domain by the hinge, CH2 and CH3
CC domains of human IgG1. The extracellular portion of the chimeric
CC receptor contains a fragment of CD4 (amino acids 1-394 or 1 200 of the
CC CD4 sequence) which specifically recognises and binds HIV-infected cells,
CC but does not mediate HIV infection. The extracellular domain of the
CC receptor is separated from the cell membrane by 48 or 72 angstroms, or by
CC one or more proteinaceous alpha-helices. The cells expressing the
CC receptor are preferably T cells, B cells, neutrophils, or dendritic
CC cells. The therapeutic cells expressing the chimeric receptor are
CC administered to a mammal to treat HIV infection.
SQ Sequence 182 AA;

Query Match 16.7%; Score 87; DB 1; Length 182;
Best Local Similarity 35.7%; Pred. No. 1.23e+01;
Matches 10; Conservative 12; Mismatches 5; Indels 1; Gaps 1;

Db 115 ISGFLFAEIVSIFVLAVGVYFIAGQDGV 142
:::::|| | ||:|: :|| ||: :|::
Qy 12 LNSYLFVPIGSIYVVDIAV-FITLDDAI 38

RESULT 12
ID R78670 standard; Protein; 182 AA.
AC R78670;
DT 11-APR-1996 (first entry)
DE T3 gamma receptor protein.
KW Chimeric receptor; CD4; T-cell receptor; HIV; cytolysis;
KW human immunodeficiency virus; adoptive immunotherapy;
KW T3 gamma receptor.
OS Homo sapiens.
FH Key Location/Qualifiers
FT region 141. .182
FT /note= "amino acids 141-182 constitute a region
FT sufficient for cytolytic signal
FT transduction"
PN WO9521528-A1.
PD 17-AUG-1995.
PF 12-JAN-1995; U00454.
PR 14-FEB-1994; US-195395.
PR 02-AUG-1994; US-284391.
PA (GEHO) GEN HOSPITAL CORP.
PI Banapour B, Kolanus W, Romeo C, Seed B;
DR WPI; 95-292893/38.
PT Target cytolysis of HIV-infected cells - by chimeric CD4
PT receptor-bearing cells
PS Example 11; Fig 17; 118pp; English.
CC Intracellular and transmembrane signal transducing domains are
CC derived from the T-cell receptor proteins CD3 delta (R78669) and
CC T3 gamma (R78670) and from the B-cell receptor proteins mb1
CC (R78671) and B29 (R78672). Each may be linked to the
CC extracellular domain of CD4 to obtain a chimeric receptor useful
CC for targeted cytolysis of HIV-infected cells.
SQ Sequence 182 AA;

Query Match 16.7%; Score 87; DB 1; Length 182;
Best Local Similarity 35.7%; Pred. No. 1.23e+01;
Matches 10; Conservative 12; Mismatches 5; Indels 1; Gaps 1;

Db 115 ISGFLFAEIVSIFVLAVGVYFIAGQDGV 142
:::::|| | ||:|: :|| ||: :|::
Qy 12 LNSYLFVPIGSIYVVDIAV-FITLDDAI 38

RESULT 13
ID W83144 standard; Protein; 182 AA.
AC W83144;
DT 03-FEB-1999 (first entry)
DE T3 gamma receptor protein.
KW Human; zeta; eta; gamma; membrane bound chimeric [illegible]
KW tumour; cancer cell; autoimmune-generated cell; T [illegible]
KW CD4; B cell receptor; Fc receptor; pathogen; bacte[illegible]
KW protozoan; viral.
OS Unidentified.
PN US5843728-A.
PD 01-DEC-1998.
PF 05-APR-1995; 417495.
PR 06-MAR-1992; US-847569.
PR 07-MAR-1991; US-665961.
PR 28-FEB-1994; US-203869.
PR 05-APR-1995; US-417495.
PA (GEHO) GEN HOSPITAL CORP.
PI Kolanus W, Romeo C, Seed B;
DR WPI; 99-044582/04.
PT Membrane-bound chimeric receptors - comprising ext[illegible]
PT which recognises and binds a target cell and an in[illegible]
PT portion of e.g. a T cell receptor
PS Example 10; Fig 17; 57pp; English.
CC The present invention describes DNA encoding a mem[illegible]
CC receptor comprising: (a) an extracellular portion [illegible]
CC recognises and binds a target cell or a target int[illegible]
CC an intracellular portion of a T-cell receptor, CD4,
CC polypeptide, a B-cell receptor polypeptide or an I[illegible]
CC The present sequence represents a T3 gamma recepto[illegible]
CC example of the present invention. Cells expressing[illegible]
CC the present invention can be administered to mamma[illegible]
CC pathogens (e.g. bacteria, fungi, protozoa or virus[illegible]
CC cancer cells or autoimmune-generated cells.
SQ Sequence 182 AA;

Query Match 16.7%; Score 87; DB 1; Leng[illegible]
Best Local Similarity 35.7%; Pred. No. 1.23e+01;
Matches 10; Conservative 12; Mismatches 5; [illegible]

Db 115 ISGFLFAEIVSIFVLAVGVYFIAGQDGV 142
:::::|| | ||:|: :|| ||: :|::
Qy 12 LNSYLFVPIGSIYVVDIAV-FITLDDAI 38

RESULT 14
ID R50727 standard; peptide; 46 AA.
AC R50727;
DT 14-MAY-1996 (first entry)
DE G-protein coupled receptor IM5 consensus polypepti[illegible]
KW G-protein coupled receptor; ligand binding assay; [illegible]
KW psychotic disorder; schizophrenia; dopamine; cAMP; [illegible]
KW muscarinic acetylcholine; adrenergic; endothelin; [illegible]
KW rhodopsin, opsin; odorant; cytomegalovirus; sec [illegible]
OS Synthetic.
PN WO9405695-A1.
PD 17-MAR-1994.
PF 09-SEP-1993; U08528.
PR 10-SEP-1992; US-943[illegible].
PA (UYNY) UNIV NEW YORK STATE.
PI Murphy RB, Schuster TJ;
DR WPI; 94-101120/12.
PT Polypeptides of G-coupled receptor proteins (GPRS)
PT binding GPR ligands or modulating GPR binding
PS Claim 9; Page 30; 160pp; English.
CC Polypeptides R50719-83 are based on the consensus [illegible]
CC sequence from G protein coupled receptor (GPR) pro[illegible]
CC in G-protein coupled receptor ligand binding assay[illegible]
CC used to identify fragments prot. transmembrane tra[illegible]
CC proteins (see R48686-R48756 for examples) which re[illegible]
CC activity such as binding a GPR ligand or modulatin[illegible]
CC a GPR (see R48759-R48798, R50569-R50807 and R50[illegible]
CC polypeptide fragments). The polypeptide fragments[illegible]
CC compositions for treating subjects suffering from[illegible]
CC a GPR abnormality eg. a psychotic disorder such a[illegible]
SQ Sequence 46 AA;

Query Match 16.1%; Score 84; DB 1; Leng[illegible]

```
  Best Local Similarity  46.2%;  Pred. No. 2.13e+01;
  Matches   12;  Conservative    9;  Mismatches   3;  Indels   2;  Gaps   2;

Db      19 IFNLMAFLFCSLGSFYI-PIAVILII 43
           || | ::|| ::|||:|:  |||::|:
Qy      10 IF-LNSYLFVPIGSIYVVDIAVFITL 34


RESULT   15
ID   W02919 standard; peptide; 46 AA.
AC   W02919;
DT   04-OCT-1996  (first entry)
DE   G-protein coupled receptor TM5 consensus polypeptide #9.
KW   G-protein coupled receptor; ligand binding assay; transmembrane domain;
KW   schizophrenia; dopamine; cAMP; adenosine; thrombin; adrenergic; opsin;
KW   muscarinic acetylcholine; endothelin; bombesin; endocrine; rhodopsin;
KW   odorant; cytomegalovirus; serotonergic.
OS   Synthetic.
PN   US5508384-A.
PD   16-APR-1996.
PF   10-SEP-1992; 943246.
PR   10-SEP-1992; US-943246.
PR   09-SEP-1993; US-118270.
PA   (UYNY ) UNIV NEW YORK STATE.
PI   Murphy RB,  Schuster DI;
DR   WPI; 96-208785/21.
PT   New dopamine receptor peptide  useful as antipsychotic agent, e.g.
PT   for treating schizophrenia
PS   Disclosure; Column 279-280; 184pp; English.
CC   Polypeptides W02747-W02910 are based on the consensus transmembrane
CC   domain III sequence from G-protein coupled receptor (GPR) proteins
CC   and can be used in GPR ligand binding assays. The assays can be used
CC   to identify fragments, pref. transmembrane fragments, from GPR proteins
CC   (see W02697-W02635) which retain biological activity such as binding a
CC   GPR ligand or modulating a GPR ligand binding to a GPR (see W02747-W02999
CC   for examples of polypeptide fragments). The polypeptide fragments can be
CC   used in compsns. for treating subjects suffering from a pathology related
CC   to a GPR abnormality e.g. schizophrenia.
SQ   Sequence   46 AA;

  Query Match           16.1%;  Score 84;  DB 1;  Length 46;
  Best Local Similarity  46.2%;  Pred. No. 2.13e+01;
  Matches   12;  Conservative    9;  Mismatches   3;  Indels   2;  Gaps   2;

Db      19 IFNLMAFLFCSLGSFYI-PIAVILII 43
           || | ::|| ::|||:|:  |||::|:
Qy      10 IF-LNSYLFVPIGSIYVVDIAVFITL 34


Search completed: Fri Aug 25 17:56:08 2000
Job time : 17 secs.
```